ARCHIVE

for

RATIONAL MECHANICS

and

ANALYSIS

Edited by

C. TRUESDELL

Volume 4, Number 1

SPRINGER-VERLAG
BERLIN-GÖTTINGEN-HEIDELBERG
(Postverlagsort Berlin · 4. 11. 1959)

Mechanicam vero duplicem Veteres constituerunt: Rationalem *quae per Demonstrationes accurate procedit, & Practicam. Ad practicam spectant Artes omnes Manuales, a quibus utique* Mechanica *nomen mutuata est. Cum autem Artifices parum accurate operari soleant, fit ut* Mechanica *omnis a* Geometria *ita distinguatur, ut quicquid accuratum sit ad* Geometriam *referatur, quicquid minus accuratum ad* Mechanicam. *Attamen errores non sunt Artis sed Artificum. Qui minus accurate operatur, imperfectior est Mechanicus, & si quis accuratissime operari posset, hic foret Mechanicus omnium perfectissimus.* NEWTON

La généralité que j'embrasse, au lieu d'éblouïr nos lumieres, nous découvrira plutôt les véritables loix de la Nature dans tout leur éclat, & on y trouvera des raisons encore plus fortes, d'en admirer la beauté & la simplicité. EULER

Ceux qui aiment l'Analyse, verront avec plaisir la Méchanique en devenir une nouvelle branche ...: LAGRANGE

The ARCHIVE FOR RATIONAL MECHANICS AND ANALYSIS nourishes the discipline of mechanics as a deductive, mathematical science in the classical tradition and promotes pure analysis, particularly in contexts of application. Its purpose is to give rapid and full publication to researches of exceptional moment, depth, and permanence.

Each memoir must meet a standard of rigor set by the best work in its field. Contributions must consist largely in original research; on occasion, an expository paper may be invited.

English, French, German, Italian, and Latin are the languages of the Archive. Authors are urged to write clearly and well, avoiding an excessively condensed or crabbed style.

Manuscripts intended for the Archive should be submitted to an appropriate member of the Editorial Board.

The ARCHIVE FOR RATIONAL MECHANICS AND ANALYSIS appears in numbers struck off as the material reaches the press; five numbers constitute a volume. Subscriptions may be entered through any agent. The price is DM 96.— per volume.

For each paper 75 offprints are provided free of charge.

Invariant and Complete Stress Functions for General Continua

C. TRUESDELL

Contents

1. Definition of stress functions. Completeness. If a particular rule enables us to find finite algebraic combinations of the derivatives of a set of arbitrary functions, possibly supplemented by quantities associated with the geometry of the space in question, such that these combinations when substituted for the stress tensor satisfy the equations of equilibrium or motion identically in the arbitrary functions, the rule is said to furnish a *solution in terms of stress functions*. Such solutions are of two kinds:

1. Solutions of particular dynamical equations, such as those of linear elasticity, classical fluid mechanics, *etc.*

2. Solutions of the dynamical equations valid for all continuous media in a particular space.

In the above statement, the "arbitrary" functions are restricted to belong to a certain class, such as functions which are four times continuously differentiable, or harmonic, or biharmonic, but the class must not be specified in terms of material properties, nor may it coincide with the class of solutions of the original equations, for otherwise the definition would be vacuous.

By their definition, stress functions provide sufficient conditions for solving the dynamical equations in question. From any such rule, by particularizing the arbitrary functions we may derive any number of special exact solutions. Thus a solution in stress functions is useful for experimentation in search of particular cases, for inverse and semi-inverse methods, *etc.* But this is not enough. We should prove also that the solution so obtained is *general*, or *complete:* Corresponding to any stress field satisfying the given equations there exists at least one suitable choice of the stress functions — in a word, that the rule expresses a necessary as well as a sufficient condition. Only on the basis of a theorem of completeness are we justified in inferring from use of the stress functions any quality of the theory itself. With such a theorem, however, we may, at least in principle, lay aside the differential equations once and for all and enjoy the advantages of explicit, finite formulae.

The second kind of stress functions is obviously of a more embracing nature than the first. However, when a solution of the second kind has been found, it does not, in general, yield a solution in stress functions for any particular material; rather, when a solution of the second kind is substituted into the relevant constitutive equations, there results, usually, a system of differential or integro-differential or functional equations to be satisfied by the stress functions. This corresponds to the very nature of constitutive equations, which single out among all dynamically possible stresses those compatible with a particular kind of material behavior.

The process just mentioned is not often an efficient one for solution in a particular theory. For example, in linear elasticity the stress functions of GALERKIN and PAPKOVICH, which are stress functions of the first kind, furnish a more practical approach to the solution of special problems and the proof of general theorems. But the moment that linear elasticity and closely related theories are left behind, these stress functions fail altogether. Stress functions of the second kind, however, are valid independently of constitutive equations. They should not be looked upon so much as tools to be applied in any special theory but as means of finding *properties common to all theories* and of *comparing the results of different theories*. The rise of various non-linear theories of materials in the past decade suggests that applications of this kind, though presently few, may soon be much demanded[1].

2. Nature and purpose of this report. Here we consider only the second kind of solution, stress functions for general continua[2], independent of constitutive assumptions. Since the ever growing literature, while giving evidence of broad and continuing interest, includes many rediscoveries, it has seemed worthwhile to put in one place all major known results and methods. The derivations are carried through in general co-ordinates, yielding results in *invariant form*, and the solutions are proved to be *complete* within a stated class.

[1] Stress functions do not furnish the only possibility of such comparisons. In fact, most of the presently known results of this kind have been obtained by SIGNORINI's theory of stress means; for an exposition, see §§ 220—222 of the forthcoming work by TRUESDELL & TOUPIN, *The classical field theories*, Handbuch der Physik III/1, Springer-Verlag.

[2] A bibliography of works on stress functions for the classical linear theory of elasticity is subjoined at the end of the report.

The equations to be solved, in the most general case envisioned, are

$$T^{km}{}_{,m} = 0, \qquad T^{km} = T^{km}, \tag{2.1}$$

where the quantities T^{km} are the contravariant components of an absolute tensor field and where the comma denotes covariant differentiation based upon an affine connection in a real space of n dimensions.

The special character of many results recently published reflects haphazard methods, unnecessary had use been made of either of the two known general approaches:

1. Representation of a solenoidal field as the curl of a vector potential.

2. Use of the principle of virtual work and its converse.

Both methods should be applicable to curved spaces. That the former method has been used only for flat spaces reflects the fact that the general theory of potentials is unknown for curved spaces. In a flat space of n dimensions, the equation $T^{k_1 \cdots k_p m}{}_{,m} = 0$ may be regarded as a system of n^p independent vector equations of the type $V^m{}_{,m} = 0$, so that the ordinary theorem of the vector potential suffices[3]. In a curved space, the equations $T^{km}{}_{,m} = 0$ are not known to be reducible to a system of equations of the form $V^m{}_{,m} = 0$, so the fact that the general solution of $V^m{}_{,m} = 0$ is known has not led to solution of our problem. The latter method, in its first stages, applies equally easily to curved spaces. To get fully explicit results by its means, however, the conditions of compatibility in the appropriate space must be exhibited. Calculation of the conditions of compatibility and proof of the theorem of the vector potential are thus seen to be closely related problems, neither of which has been solved for a general curved space. For ordinary curved surfaces, however, enough is known that we can illustrate the power of the latter method by obtaining the most general equilibrated stresses in several classes of curved membranes.

Both methods, in contrast to the more popular method of hit-or-miss trials, yield or imply proofs of completeness.

3. Mechanical situations to which the results may be applied. In an inertial co-ordinate system in a real Euclidean space of three dimensions, the stress tensor t^{km} satisfies CAUCHY's laws of motion

$$\varrho \ddot{x}^k = t^{km}{}_{,m} + \varrho f^k, \qquad t^{km} = t^{mk}. \tag{3.1}$$

These laws express the local balance of linear momentum and of moment of momentum, respectively, on the supposition that there are neither extrinsic couples nor couples regarded as representing the actions which maintain material continuity. The tensor $\boldsymbol{t}$ is presumed to include any electromagnetic stresses that may be present. ϱ is the density, $\ddot{\boldsymbol{x}}$ the acceleration, and $\boldsymbol{f}$ the extrinsic force per unit mass.

CAUCHY's laws (3.1) assume the form (2.1) in various special cases.

Case 1. *Internally free three-dimensional medium in equilibrium.* If $\dot{\boldsymbol{x}}$, the velocity, is constant and uniform, and if $\boldsymbol{f} = 0$, (3.1) assumes the form (2.1) with $\boldsymbol{T} = \boldsymbol{t}$.

[3] *Cf.* FINZI & PASTORI [1949, 1, Cap. IV, § 7].

Case 2. *Equilibrium of a three-dimensional medium subject to conservative extrinsic force.* If for some scalar $U(\boldsymbol{x}, t)$ we have

$$\varrho f_k = -U_{,k}, \tag{3.2}$$

and again $\dot{\boldsymbol{x}} = \text{const.}$, (3.1) assumes the form (2.1) if we put $\boldsymbol{T} = \boldsymbol{t} - U\boldsymbol{1}$.

In any case, if the general solution of ${t^{km}}_{,m} = 0$ is known, all that is needed in order to find the general solution of ${t^{km}}_{,m} = -\varrho f^k$ is to add a particular integral. In rectangular Cartesian co-ordinates, such an integral ${}_{\mathrm{P}}t^{km}$ is furnished for convex domains by

$$||{}_{\mathrm{P}}t^{km}|| = \begin{Vmatrix} -\int \varrho f_x \, dx & 0 & 0 \\ 0 & -\int \varrho f_y \, dy & 0 \\ 0 & 0 & -\int \varrho f_z \, dz \end{Vmatrix}. \tag{3.3}$$

In curved spaces, it may be more difficult to find a particular integral. For a Riemannian space, the search for a particular integral may be systematized as follows[4]: Let u^k be any contravariant vector satisfying the partial differential system

$$u^k{}_{,m}{}^{,m} - R^k_m u^m + \varrho (f^k - \ddot{x}^k) = 0, \tag{3.3a}$$

where R^k_m is the contracted curvature tensor; then a particular integral ${}_{\mathrm{P}}t^{km}$ of the system (3.1) is given by

$${}_{\mathrm{P}}t^{km} = 2u^{(k,m)} - u^p{}_{,p}\, g^{km}. \tag{3.3b}$$

The problem of finding a particular integral of the underdetermined tensorial equations is thus reduced to that of finding a particular integral of a determinate vectorial system of Helmholtz type: $\nabla^2 \boldsymbol{u} = f(\boldsymbol{x})\, \boldsymbol{u} + \boldsymbol{g}(\boldsymbol{x})$.

Case 3. *Steady motion of a three-dimensional medium.* In virtue of EULER's equation of continuity,

$$\frac{\partial \varrho}{\partial t} + (\varrho \dot{x}^k)_{,k} = 0, \tag{3.4}$$

we may put CAUCHY's first law $(3.1)_1$ into the form

$$\frac{\partial (\varrho \dot{x}^k)}{\partial t} = (t^{km} - \varrho \dot{x}^k \dot{x}^m)_{,k} + \varrho f^k. \tag{3.5}$$

In the case of steady motion subject to no extrinsic force, (3.5) reduces to $(2.1)_1$ if we put $\boldsymbol{T} = \boldsymbol{t} - \varrho \dot{\boldsymbol{x}} \dot{\boldsymbol{x}}$, and $(2.1)_2$ is satisfied in virtue of $(3.1)_2$.

Cases 2 and 3 may be combined.

Case 4. *Unsteady motion of a three-dimensional medium.* Let indices Γ, Δ run from 1 to 4, and retain k, m as indices with range 1 to 3. Put

$$x^\Gamma \equiv (x^k, t), \qquad V^\Gamma \equiv (\dot{x}^k, 1), \tag{3.6}$$

and

$$||t^{\Gamma\Delta}|| \equiv \begin{Vmatrix} t^{km} & 0 \\ 0 & 0 \end{Vmatrix}, \tag{3.7}$$

and define a *stress-momentum tensor* $\boldsymbol{T}$ as follows:

$$T^{\Gamma\Delta} \equiv t^{\Gamma\Delta} - \varrho V^\Gamma V^\Delta. \tag{3.8}$$

[4] For a flat space, the method is due to SCHAEFER [1953, 5, § 5].

Thus, *e.g.*, $T^{r4}=-\varrho V^r=T^{4r}$, $T^{44}=-\varrho$. Suppose that $\boldsymbol{f}=0$, and select *rectangular Cartesian space co-ordinates* x^k. Then (3.4) and (3.5) together assume the form $(2.1)_1$, where the comma now stands for partial differentiation. Thus the laws of motion in an inertial frame assume the form to which (2.1) would reduce *if space-time were flat*. This observation suffices for our later analysis of unsteady motion, which does not presume any geometrical structure of space-time[5].

Case 5. *Plane motion subject to plane stress.* When $\dot{\boldsymbol{x}}$ and $\boldsymbol{t}$ are plane, all the foregoing cases become applicable, but the dimension of the space drops by one. In Case 4, for example, the equations of plane unsteady motion assume the same form, in an inertial frame, as the equations of equilibrium in a three-dimensional flat space.

The equations of equilibrium of a shell having first and second fundamental forms $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively, are of the form[6]

$$\begin{aligned} &S^{\delta}{}_{,\delta}+b_{\gamma\delta}S^{\gamma\delta}+F=0,\\ &S^{\gamma\delta}{}_{,\delta}-a^{\gamma\delta}b_{\sigma\delta}S^{\delta}+F^{\delta}=0,\\ &M^{\delta}{}_{,\delta}+b_{\gamma\delta}M^{\gamma\delta}+e_{\gamma\delta}S^{\gamma\delta}+L=0,\\ &M^{\gamma\delta}{}_{,\delta}-a^{\gamma\sigma}b_{\sigma\delta}M^{\delta}+a^{\gamma\delta}e_{\delta\sigma}S^{\sigma}+L^{\gamma}=0, \end{aligned} \tag{3.9}$$

where the comma denotes covariant differentiation based upon the surface metric $\boldsymbol{a}$, where the indices have the range 1, 2, and where $e_{\gamma\delta}\equiv+\sqrt{a}\,\varepsilon_{\gamma\delta}$, $\varepsilon_{\gamma\delta}$ being the permutation symbol such that $\varepsilon_{12}=+1$. F^{δ} and F are the tangential and normal extrinsic forces, L^{δ} and L the corresponding couples; $S^{\gamma\delta}$ and S^{δ} are the membrane stress and cross-force resultants, $M^{\gamma\delta}$ and M^{δ} the corresponding couples. Only in special cases do (3.9) reduce to the form (2.1).

Case 6. *Equilibrium of an internally free membrane.* In all classical studies of shells, it is assumed that $L=0$ and $M^{\delta}=0$. If in addition we have $S^{\delta}=0$, $L^{\gamma}=0$, and $M^{\gamma\delta}=0$, $(3.9)_4$ is satisfied identically, $(3.9)_2$ and $(3.9)_3$ reduce to

$$S^{\gamma\delta}{}_{,\delta}+F^{\delta}=0, \qquad S^{\gamma\delta}=S^{\delta\gamma}, \tag{3.10}$$

while $(3.9)_1$ becomes an algebraic equation relating $S^{\gamma\delta}$ and F:

$$b_{\gamma\delta}S^{\gamma\delta}+F=0. \tag{3.11}$$

A tensor $\boldsymbol{S}$ satisfying (3.10) and (3.11) is said to describe a *state of membrane stress* upon the surface whose fundamental forms are $\boldsymbol{a}$ and $\boldsymbol{b}$. When $F^{\delta}=0$, (3.10) is of the form (2.1). If (3.10) is solved, substitution of the result into

[5] However, it would be permissible to use an invariant formulation throughout, since results of TOUPIN show that the laws of classical mechanics may be expressed as tensorial equations in an affine space-time such that the affine connection vanishes in all inertial frames. See § 4 of his *World invariant kinematics*, Arch. Rational Mech. Anal. **1**, 181–211 (1958). We emphasize that neither his treatment nor ours presupposes a four-dimensional metric or makes any commitment in regard to relativity.

[6] See, *e.g.*, § 26 of J. L. ERICKSEN & C. TRUESDELL: *Exact theory of stress and strain in rods and shells*. Arch. Rational Mech. Anal. **1**, 295–323 (1958).

(3.11) yields a differential equation to be satisfied by the stress function or potential. Solution of (3.10) by itself constitutes an intrinsic problem; when (3.11) is taken into account as well, the problem is no longer an intrinsic one.

Part I. Flat spaces

4. General solution for a flat space of n dimensions. In order that $T^{km}{}_{,m}=0$ in a Euclidean space of any number of dimensions, application of the classical theorem of the vector potential shows it to be necessary and sufficient that[7]

$$T^{km}=b^{kmp}{}_{,p}, \quad \text{where} \quad b^{kmp}=-b^{kpm}. \tag{4.1}$$

The condition $T^{km}=T^{mk}$ may now be written in the form

$$(b^{kmp}-b^{mkp})_{,p}=0. \tag{4.2}$$

This condition, in turn, is equivalent to the existence of a tensor $\boldsymbol{c}$ such that

$$b^{kmp}-b^{mkp}=c^{kmpq}{}_{,q}, \tag{4.3}$$

where

$$c^{kmpq}=-c^{kmqp}=-c^{mkpq}. \tag{4.4}$$

Therefore

$$2b^{kmp}=(c^{pkmq}+c^{mpqk}+c^{kmpq})_{,q}, \tag{4.5}$$

so that (3.1) becomes

$$T^{km}=h^{pkmq}{}_{,qp}, \tag{4.6}$$

where

$$\begin{aligned} h^{pkmq}&\equiv\tfrac{1}{2}(c^{pkmq}+c^{mqpk}),\\ h^{pkmq}&=-h^{kpmq}=-h^{pkqm}=h^{mqpk}. \end{aligned} \tag{4.7}$$

The elegant foregoing derivation, given by DORN & SCHILD[8], shows that (4.6) furnishes the ***general solution of Cauchy's laws*** for equilibrium of an internally free body in a flat space of any dimension.

---

[7] As observed by GWYTHER [1913], for a stress tensor which is not symmetric the analysis breaks off at this point. In a more general system of mechanics initiated by VOIGT and developed by E. & F. COSSERAT, not only is the stress tensor not generally symmetric, but also there is a couple stress tensor in addition to the usual one. In this system of mechanics, all equations of equilibrium are prescriptions of divergences. Stress functions may be introduced by the procedure given above; *cf.* GÜNTHER [1958, 2, § 3].

[8] [1956, 2]. The basic idea was suggested by BELTRAMI [1892, 2]. *Cf.* also MORERA [1892, 3].

A variant of this derivation had been given earlier by GÜNTHER [1954, 1, § 1]. He begins by introducing the skew-symmetric dual tensor of fourth order,

$$(*) \qquad T_{kmpq}\equiv e_{kmn}e_{pqs}T^{ns},$$

which he interprets as a "transversal stress tensor". An explicit solution for T_{kmpq} in terms of stress functions is simply obtained. In DORN & SCHILD's proof, the duals appear in (5.1). While GÜNTHER's solution for the *dual* tensor is indeed valid, as he says, in a flat space of any dimension, to derive from it the ordinary stress tensor we must use the inverse of (*) and hence presume the number of dimensions to be three. Of course GÜNTHER's proof can be adjusted to the n-dimensional case also, but DORN & SCHILD's proof is equally simple and natural in all cases.

To infer (4.7), the theorem of the vector potential has been applied twice, *viz.*, to (4.1) and to (4.2). From writings on the theory of the potential[9] may be inferred the following ***theorem of completeness***: Let $\mathscr{D}$ be a regular region[10] with boundary $\mathscr{B}$. *In $\mathscr{D}$, let the components T^{km} be twice continuously differentiable functions satisfying* (2.1), *and in $\mathscr{D}+\mathscr{B}$, once continuously differentiable. If $\mathscr{D}$ is infinite, assume that $T^{km}=O(r^{-2-\delta})$ and $T^{km}{}_{,p}=O(r^{-3-\delta})$, $\delta>0$. Then there exist infinitely many tensors h^{pkmq} such that T^{km} may be expressed in the form* (4.6).

The conditions stated are merely sufficient for completeness and are by no means the weakest possible[11].

From (4.6) it is clear that the tensor h^{pkmq} is indeterminate to within a tensor ${}_0h^{pkmq}$ such that ${}_0h^{pkmq}{}_{,qp}=0$. We shall not take up the general problem of finding conditions sufficient to individualize a particular tensor within this class, though we shall allude to special cases below.

5. Three-dimensional flat space. In the three-dimensional case, set

$$a_{rs}\equiv\tfrac{1}{4}e_{rpk}\,e_{sqm}\,h^{pkmq}, \tag{5.1}$$

so that

$$h^{pkmq}=e^{krp}\,e^{msq}\,a_{rs}, \qquad a_{rs}=a_{sr}. \tag{5.2}$$

Then (4.6) becomes the ***Gwyther-Finzi general solution***[12]:

$$T^{km}=e^{krp}\,e^{msq}\,a_{rs,pq}. \tag{5.3}$$

---

[9] The result is not to be read off from any work I have seen, but a proof may be constructed by use of the apparatus assembled to prove the ordinary theorem of the vector potential. *Cf., e.g.,* § 83 of R. S. PHILLIPS: *Vector Analysis*, New York, Wiley 1933 (the method used in § 49 yields only local completeness). More detailed consideration is given to the problem by L. LICHTENSTEIN, §§ 12—14 of Chapter III of *Grundlagen der Hydromechanik*, Berlin, Springer, 1929. In these treatments, the potentials are rendered unique by requiring them to be solenoidal, but this is not convenient here.

[10] Regular regions are defined in §§ 8—9 of Ch. IV of O. D. KELLOGG: *Foundations of Potential Theory*, Berlin, Springer, 1929. Such regions need not be bounded or simply connected, and their bounding surfaces may have corners and edges, subject to specified restrictions.

[11] I do not attempt to obtain in the present case a weakening of hypotheses so as to allow certain discontinuities in the components T^{km} and their derivatives corresponding to LICHTENSTEIN's treatment of the ordinary theorem of the vector potential, because it seems preferable to base a more general analysis on the concepts of H. WEYL: *The method of orthogonal projection in potential theory*, Duke Math. J. **7**, 411—444 (1940), but this requires a whole new program from the start. Furthermore, general discontinuities in T^{km} are not of interest, since the integral conservation laws to which (2.1) are equivalent for continuously differentiable fields T^{km} imply restrictions upon the discontinuities allowable. A truly general treatment should be based upon the integral conservation laws throughout.

[12] GWYTHER [1912] obtained (5.3) in orthogonal curvilinear co-ordinates, writing out the special cases appropriate to rectangular Cartesian, cylindrical polar, and spherical polar co-ordinates (*cf.* also [1911]). His steps are such as to imply the necessity of the result; the sufficiency is immediate. B. FINZI [1934, 2, § 3] observed that (5.3) yields a symmetric tensor satisfying $T^{km}{}_{,m}=0$; for a proof of completeness he was content to refer to the previously known fact that the special cases (5.5) and (5.6) are complete. The Cartesian tensor form of (5.3) is almost evident from an equation of KLEIN & WIEGHARDT [1905, Eq. (33)], but they did not obtain it.

FINZI noticed that the tensor $\boldsymbol{a}$ is indeterminate[13] to within an arbitrary symmetric tensor ${}_0\boldsymbol{a}$ satisfying $e^{krp}\, e^{msq}\, {}_0a_{rs,pq} = 0$. From a known result concerning the conditions of compatibility, such a tensor is of the form ${}_0a_{mr} = b_{(m,r)}$, where $\boldsymbol{b}$ is an arbitrary vector. For a given $\boldsymbol{a}$, in rectangular Cartesian co-ordinates, we may choose $\boldsymbol{b}$ so as to satisfy one or the other of the conditions

$$\begin{aligned} b_{(m,r)} &= -a_{mr}, \quad r \neq m \\ b_{m,m} &= -a_{mm} \quad \text{(unsummed)}. \end{aligned} \tag{5.4}$$

These two choices of $\boldsymbol{b}$ show that for a particular rectangular Cartesian co-ordinate system there is no loss in generality in asuming in the first place that $\boldsymbol{a}$ is diagonal, or that the diagonal components of $\boldsymbol{a}$ are zero. The former alternative yields

$$T_{xx} = a^2_{,zz} + a^3_{,yy}, \quad T_{xy} = -a^3_{,yx}, \quad \text{etc.}, \tag{5.5}$$

with $a^1 \equiv a_{xx}$, $a^2 \equiv a_{yy}$, $a^3 \equiv a_{zz}$; the latter alternative yields

$$T_{xx} = -2a^4_{,yz}, \quad T_{xy} = (a^4_{,x} + a^5_{,y} - a^6_{,z})_{,z}, \quad \text{etc.}, \tag{5.6}$$

with $a^4 \equiv a_{23}$, $a^5 \equiv a_{31}$, $a^6 \equiv a_{12}$. These two forms of the general solution were obtained by MAXWELL and MORERA[14], respectively. As follows from the special choices of $\boldsymbol{a}$ made to derive them, these special forms are not invariant under transformations even of rectangular Cartesian co-ordinates. The explicit form for (5.3), with no restrictions on the six potentials, may be obtained by adding together the right-hand sides of (5.5) and (5.6). Other special choices of the potentials are possible[15], but it by no means follows that a solution obtained by imposing three arbitrary conditions on the six potentials a_{km} remains complete[16].

[13] This fact is used by PERETTI [1949, 2] to show that it is possible to choose $\boldsymbol{a}$ in such a way that from its components may be obtained simple expressions for the resultant force and moment of the stresses on a surface. *Cf.* also BLOKH [1950, 1]. GÜNTHER [1954, 1, § 3] gives simple expressions for the resultant force and torque on a body in terms of integrals of stress functions around particular curves. SCHAEFER [1955, 3] discusses the nature of ${}_0\boldsymbol{a}$.

[14] [1868] [1869]; [1892, 1]. Using results given by BELTRAMI [1892, 2], MORERA [1892, 3] modified his derivation so as to yield (5.5) alternatively to (5.6). *Cf.* also GWYTHER [1913]. A literature devoted mainly to rediscovery of known results regarding this subject has arisen recently; *cf.* KUZMIN [1945], WEBER [1948], MORINAGA & NÔNO [1950, 3], SCHAEFER [1953, 5], LANGHAAR & STIPPES [1954, 1], ORNSTEIN [1954, 2].

[15] *Cf.* MORINAGA & NÔNO [1950, 3, § 3]. BLOKH [1950, 1] lists the essentially different forms that result from such special choices: 5 in rectangular Cartesian co-ordinates, 20 in general co-ordinates, 18 in cylindrical co-ordinates, 10 in cylindrical co-ordinates for rotationally symmetric problems, 19 in spherical co-ordinates.

[16] What reductions are possible is not obvious. In writings on stress functions there is a deplorable custom of inferring completeness by merely counting the number of arbitrary functions. Apart from the logical gap in such inference, its danger is illustrated by the solution written down without proof of completeness by PRATELLI [1953, 4]:

$$T^{km} = e^{kpq}\, e^{mrs} (F_{,pr}\, g_{qs} + H_{,ps}\, g_{qr} + K_{,qs}\, g_{pr}). \tag{A}$$

While indeed a solution for any choice of the three arbitrary potentials F, H, K, it is not general. In fact, by using the expression for the product $\varepsilon^{kpq}\, \varepsilon^{mrs}$ as a determinant of δ^t_n's, we may put (A) into the form

$$T^{km} = P^{,q}{}_{,q}\, g^{km} - P^{,km}, \tag{B}$$

where $P \equiv F - H + K$, and it is easy to exhibit solutions of $T^{km}{}_{,m} = 0$ which cannot be expressed in the form (B).

While AIRY's stress function for steady plane motion subject to plane stress is formally a special case of the foregoing, a treatment which is fully two-dimensional from the start is preferable. I do not develop it here because the existing literature is adequate[17].

6. Application to the axially symmetric case. If we write out (5.3) explicitly in cylindrical polar co-ordinates, at the same time supposing that all derivatives with respect to the azimuth angle are zero, for the physical components $\widehat{ik}$ of $\boldsymbol{T}$ we obtain[18]

$$\begin{aligned}
\widehat{rr} &= a^2_{,zz} + \frac{1}{r} a^3_{,r} - \frac{2}{r} a^5_{,z}, \\
\widehat{\vartheta\vartheta} &= a^3_{,rr} + a^2_{,zz} - 2a^5_{,rz}, \\
\widehat{zz} &= a^2_{,rr} + \frac{2}{r} a^2_{,r} - \frac{1}{r} a^1_{,r}, \\
\widehat{rz} &= -\left[a^2_{,r} + \frac{1}{r} a^2 - \frac{1}{r} a^1\right]_{,z}, \\
\widehat{r\vartheta} &= \left[a^4_{,r} - \frac{1}{r} a^4 - a^6_{,z}\right]_{,z}, \\
\widehat{\vartheta z} &= -a^4_{,rr} - \frac{1}{r} a^4_{,r} + \frac{1}{r^2} a^4 + a^6_{,zr} + \frac{2}{r} a^6_{,z},
\end{aligned} \tag{6.1}$$

where commas denote partial derivatives, and where we have set $a^1 \equiv a_{rr}$, $a^2 \equiv a_{\vartheta\vartheta}/r$, $a^3 \equiv a_{zz}$, $a^4 \equiv a_{\vartheta z}/r$, $a^5 \equiv a_{zr}$, $a^6 \equiv a_{r\vartheta}/r$. The potentials a^1, a^2, a^3, and a^5 occur only in the first four members of (6.1); the potentials a^4 and a^6, only in the last two. Axially symmetric stress distributions in which $\widehat{r\vartheta} = 0$, $\widehat{\vartheta z} = 0$ are often called *torsionless*; the most general stress of this kind is obtained by setting $a^4 = 0$, $a^6 = 0$ in (6.1). In any case, the six potentials may be reduced to three in a variety of ways[19]. For example, if we set

$$\begin{aligned}
L_{,r} &\equiv a^2_{,r} + \frac{1}{r} a^2 - \frac{1}{r} a^1, \\
M &\equiv a^2_{,zz} + \frac{1}{r} a^3_{,r} - \frac{2}{r} a^5_{,z} - L_{,zz},
\end{aligned} \tag{6.2}$$

---

[17] Discovery: AIRY [1863].

Assertion of completeness: MAXWELL [1870, pp. 192—193].

First fully satisfactory treatment, including intrinsic forms and multi-valuedness in multiply connected regions: MICHELL [1900, 1].

Adjustment to steady flow subject to hydrostatic pressure: E. R. NEUMANN [1907, 1].

Interpretation in terms of resultant torque: PHILLIPS [1934, 4], SOBRERO [1935].

General curvilinear co-ordinates: B. FINZI [1934, 2, § 1].

Energetic interpretation: L. FINZI [1956, 3, § 6].

Other references: BRAHTZ [1934, 1], BATEMAN [1936] [1938], CROCCO [1950, 2].

The completeness theorem for AIRY's function does not follow as a special case of our theorem in § 4, since a plane stress field regarded from a three-dimensional viewpoint does not vanish at ∞. Completeness is easily established directly, however.

SCHAEFER [1956, 4] approaches the plane problem as the limit of a three-dimensional one. There result, in addition to AIRY's, two stress functions representing surface loads.

[18] BRDIČKA [1957, 1, § 6]. A more symmetrical expression is given by MARGUERRE [1955, 2], but his potentials are connected by a condition of compatibility, and there is no proof of completeness.

[19] *Cf.* BLOKH [1950, 1].

then the first four members of (6.1) become[20]

$$\widehat{rr} = L_{,rr} + M, \qquad \widehat{\vartheta\vartheta} = (rM)_{,r} + L_{,zz},$$
$$\widehat{zz} = L_{,rr} + \frac{1}{r} L_{,r}, \qquad \widehat{rz} = -L_{,zr}. \tag{6.3}$$

The function L is LOVE's *stress function for torsionless axially symmetric stress fields*. Similarly, if we put

$$W \equiv -r^2 \left| a^4_{,r} - \frac{1}{r} a^4 - a^6_{,z} \right|, \tag{6.4}$$

the last two members of (6.1) become[21]

$$\widehat{r\vartheta} = -\frac{1}{r^2} W_{,z}, \qquad \widehat{\vartheta z} = \frac{1}{r^2} W_{,r}. \tag{6.5}$$

The function W is VOIGT's *stress function for purely torsional stress fields*.

7. Application to unsteady motion subject to plane stress. We now employ the device explained under Case 4 in § 3, which rests upon the observation that the equations of plane unsteady motion subject to plane stress can be cast into the form of the equations of equilibrium of a three-dimensional body, subject to a stress T^{km} related to the plane stress $t^{\alpha\beta}$ as follows: $T^{\alpha\beta} = t^{\alpha\beta} - \varrho \dot{x}^\alpha \dot{x}^\beta$, $T^{3\alpha} = T^{\alpha 3} = -\varrho \dot{x}^\alpha$, $T^{33} = -\varrho$, where Greek indices have the range 1, 2. This identification is possible in rectangular Cartesian inertial co-ordinate systems. The solution (5.3) is then valid for $\boldsymbol{T}$ and, in these co-ordinates, is general. In writing out the result, we distinguish explicitly all time differentiations and time components. The result, derived in a special co-ordinate system, turns out to be an equation having tensorial form under general time-independent co-ordinate transformations in the plane, *viz.*

$$\begin{aligned} -\varrho &= T^{33} = e^{\delta\sigma} e^{\varphi\psi} a_{\delta\varphi,\sigma\psi}, \\ -\varrho \dot{x}^\gamma &= T^{\gamma 3} = e^{\gamma\delta} e^{\varphi\psi} (a'_{\delta\varphi,\psi} - a_{3\varphi,\delta\psi}), \\ t^{\gamma\lambda} - \varrho \dot{x}^\gamma \dot{x}^\lambda &= T^{\gamma\lambda} = e^{\gamma\varphi} e^{\lambda\psi} (a_{33,\varphi\psi} + a''_{\varphi\psi}) - (e^{\gamma\varphi} e^{\lambda\psi} + e^{\gamma\psi} e^{\lambda\varphi}) a'_{3\psi,\varphi}, \end{aligned} \tag{7.1}$$

where a prime denotes $\partial/\partial t$. In the steady case, this reduces to AIRY's solution, $-a_{33}$ being AIRY's function A. The six potentials may be reduced to three in various ways. For example, if we choose $\boldsymbol{a}$ to be diagonal in a particular rectangular Cartesian co-ordinate system, we obtain[22]

$$\begin{gathered} -\varrho = a^1_{,yy} + a^2_{,xx}, \qquad \varrho \dot{x} = a^2_{,xt}, \qquad \varrho \dot{y} = a^1_{,yt}, \\ t_{xx} - \varrho \dot{x}^2 = -A_{,yy} + a^2_{,tt}, \qquad t_{yy} - \varrho \dot{y}^2 = -A_{,xx} + a^1_{,tt}, \\ t_{xy} - \varrho \dot{x} \dot{y} = A_{,xy} \end{gathered} \tag{7.2}$$

generalizing AIRY's solution to the case of arbitrary plane motion.

[20] This solution, whose completeness is easy to prove also directly from the equations of equilibrium, was obtained by LOVE [1906, § 188]. Variants are given by BRDIČKA [1957, 1, § 4].

[21] An equivalent result in rectangular Cartesian co-ordinates was given by VOIGT [1883, pp. 297—298]. *Cf.* also MICHELL [1900, 2, p. 133], PHILLIPS [1934, 3].

[22] A similar but not obviously identical solution is obtained by KILCHEVSKI [1953, 1].

There is literature[23] concerning the lineal case, when $t_{xy}=0$, $\dot{y}=0$, and t_{yy} does not appear in the basic equations; to obtain the results, put $A=a^1=0$.

8. Unsteady motion in three dimensions. For a Euclidean space of n dimensions, we have the solution (4.6), the completeness of which has been established. When $n=4$, letting Greek capital indices have the range 1, 2, 3, 4, we may put

$$A_{\Sigma\Lambda\Phi\Psi} \equiv \tfrac{1}{4} e_{\Omega\Gamma\Sigma\Lambda} e_{\Lambda\Xi\Phi\Psi} h^{\Omega\Gamma\Lambda\Xi}, \tag{8.1}$$

so that

$$h^{\Gamma\Lambda\Delta\Xi} = e^{\Lambda\Gamma\Sigma\Phi} e^{\Delta\Xi\Psi\Omega} A_{\Sigma\Phi\Psi\Omega}. \tag{8.2}$$

From (4.6) it follows that[24]

$$T^{\Gamma\Delta} = e^{\Gamma\Lambda\Sigma\Phi} e^{\Delta\Xi\Psi\Omega} A_{\Sigma\Phi\Psi\Omega,\Lambda\Xi}, \tag{8.3}$$

where (4.7) implies the following conditions of symmetry for $\boldsymbol{A}$:

$$\begin{gathered} A_{\Sigma\Phi\Psi\Omega} = -A_{\Phi\Sigma\Psi\Omega} = -A_{\Sigma\Phi\Omega\Psi}, \\ A_{\Sigma\Phi\Psi\Omega} = A_{\Psi\Omega\Sigma\Phi}. \end{gathered} \tag{8.4}$$

Inspection of (8.3) shows that only the second of these sets of symmetry conditions is essential; the first set may be abandoned without impairing the solution. The tensor $\boldsymbol{A}$ is indeterminate to within a tensor ${}_0\boldsymbol{A}$ such that $e^{\Gamma\Lambda\Sigma\Phi} e^{\Delta\Xi\Psi\Omega}\, {}_0A_{\Sigma\Phi\Psi\Omega,\Lambda\Sigma}$. The most general such tensor is a linear combination of tensors of the type $B_{\Sigma\Phi\Psi,\Omega}$; to satisfy the essential symmetry condition $(8.4)_3$ we may choose $B_{\Sigma\Phi\Psi,\Omega}+B_{\Psi\Omega\Sigma,\Phi}=0$; if, finally, we choose to satisfy also the conditions $(8.4)_{1,2}$, we have

$$\begin{aligned} A^0_{\Sigma\Phi\Psi\Omega} &= 4B_{[\Sigma\Phi][\Psi,\Omega]} + 4B_{[\Psi\Omega][\Sigma,\Phi]}, \\ &= B_{\Sigma\Phi\Psi,\Omega} - B_{\Sigma\Phi\Omega,\Psi} - B_{\Phi\Sigma\Psi,\Omega} + B_{\Phi\Sigma\Omega,\Psi} + \\ &\quad + B_{\Psi\Omega\Sigma,\Phi} - B_{\Psi\Omega\Phi,\Sigma} - B_{\Omega\Psi\Sigma,\Phi} + B_{\Omega\Psi\Phi,\Sigma}. \end{aligned} \tag{8.5}$$

To adapt this general solution for equilibrium in a flat four-dimensional space to unsteady motion in three-dimensional space, we proceed just as in § 7, except that we use (3.6)—(3.8). Writing explicitly all time components and time differentions, we find that our result, derived in a special co-ordinate system, turns out to be in tensorial form under time-independent transformations of the space co-ordinates alone. These formulae, valid for all inertial curvilinear

---

[23] The first step was taken by Euler [1757, §§ 48—49]; the result was worked out by W. Kirchhoff [1930, 2, § 1] after being suggested by E. R. Neumann [1907, § 6]; it has been rediscovered by McVittie [1953, 3, § 3].

[24] B. Finzi [1934, 2, § 5] wrote down (8.3) and symmetry conditions consisting in (8.4) and a further requirement which I do not verify; he was content to infer completeness by counting the number of assignable arbitrary functions. A somewhat involved proof was given by Morinaga & Nôno [1950, 3, § 4]; I do not follow the argument whereby they claim [*ibid.* § 5] to establish the alternative form

$$T^{\Gamma\Delta} = e^{\Gamma\Lambda\Phi\Psi} e^{\Delta\Xi\Sigma}{}_{\Psi} A_{\Phi\Xi,\Lambda\Sigma};$$

they give corresponding results in n dimensions.

co-ordinate systems, are[25]

$$\begin{aligned} -\varrho &= T^{44} = e^{smn}\, e^{pqr}\, A_{mnqr,sp}, \\ -\varrho\, \dot{x}^k &= T^{k4} = -\,e^{ksm}\, e^{pqr} (A'_{smqr,p} + 2A_{4mqr,sp}), \\ t^{km} - \varrho\, \dot{x}^k \dot{x}^m &= T^{km} = 4 e^{kpq}\, e^{msn}\, A_{4p4s,qn} + \\ &\quad + 2(e^{kpq}\, e^{msn} + e^{ksn}\, e^{mpq})\, A'_{4npq,s} + e^{kpq}\, e^{msn}\, A''_{pqsn}, \end{aligned} \tag{8.6}$$

where a prime denotes $\partial/\partial t$, and where the symmetries of the tensors A_{pqmn}, A_{4npq}, and A_{4p4q} may be read off from (8.4). This is ***Finzi's general solution*** of the equations of balance of mass and momentum in an inertial frame. In the case of equilibrium, this solution reduces to (5.3) with $4A_{4p4n} = a_{pm}$.

In a particular co-ordinate system, by means of (8.5) we may impose 14 conditions upon the 20 independent potentials occurring in the solution (8.6). For example, we are tempted to take A_{4p4m} as diagonal, A_{4mqr} as zero, and A_{mnqr} as zero when $m \neq q$ and $n \neq r$. I have been unable to prove that it is possible to choose $\boldsymbol{B}$ in (8.5) in such a way as to justify this special choice of $\boldsymbol{A}$; if it is legitimate, then, writing $A^1 \equiv 4A_{4141}, \ldots, A^4 \equiv 2A_{2323}, \ldots$, in this case we reduce (8.3) to the form

$$\begin{aligned} -\varrho &= A^4_{,xx} + A^5_{,yy} + A^6_{,zz}, \qquad \varrho\, \dot{x} = A^4_{,xt}, \ldots \\ t_{xx} - \varrho\, \dot{x}^2 &= A^2_{,zz} + A^3_{,yy} + A^4_{,tt}, \ldots, \qquad t_{xy} - \varrho\, \dot{x}\, \dot{y} = -A^3_{,xy}, \ldots, \end{aligned} \tag{8.7}$$

extending MAXWELL's formulae (8.5) so as to yield a simple yet general solution in terms of six potentials.

9. Status of the problem for flat spaces. The foregoing sections show that the *general* problem is solved in all detail for flat spaces of any dimension. It is to be hoped that this exposition may save the labor of rediscovery for those interested in the subject, enabling them to proceed at once to applications.

When theories of materials were first proposed in the eigtheenth century, solutions in arbitrary functions such as those presented above were sought earnestly, but, for the most part, sought in vain. In the nineteenth century, researches on partial differential equations turned away from such general solutions so as to concentrate upon boundary-value problems. When, in the twentieth century, the general solutions were at last obtained, slight attention was paid to them, and to this day they remain virtually unknown. Though scant use has been made of them so far, they may be enlightening for studies of underdetermined systems, where the conventional viewpoint of partial differential equations has gained little.

[25] B. FINZI [1934, 2, § 6]. In rectangular Cartesian co-ordinates, this result is rediscovered by ARZHANIKH [1952, 1] (while he uses 21 potentials, obviously one may be eliminated). KILCHEVSKI [1953, 1] observes that if $R^{\Gamma}{}_{\Delta}$ is the Ricci tensor based on the Riemannian metric tensor $G_{\Gamma\Delta}$, then in any Riemannian 4-space the quantities $T^{\Gamma\Delta} = R^{\Gamma\Delta} - \frac{1}{2} G^{\Gamma\Delta} R^{\Phi}{}_{\Phi}$ satisfy $T^{\Gamma\Delta}{}_{,\Delta} = 0$. Putting $G_{\Gamma\Delta} = \delta_{\Gamma\Delta} + \varepsilon H_{\Gamma\Delta}$, he calculates $T^{\Gamma\Delta} = \varepsilon Q^{\Gamma\Delta} + O(\varepsilon^2)$; hence follows $\partial Q^{\Gamma\Delta}/\partial x^{\Delta} = 0$, so that $Q^{\Gamma\Delta}$, in rectangular Cartesian co-ordinates, gives a solution of the type presented in the text above. A similar approach, involving a detour through relativity theory, is presented by MCVITTIE [1953, 3, § 2]; that his solution is not general is remarked by WHITHAM [1954, 3], who obtains what appears to be a special case of FINZI's solution in a special co-ordinate system. Rediscoveries of other special cases are made by MILNE-THOMSON [1957, 3] and by BLANKFIELD & MCVITTIE [1959, 1 & 2].

Application to the simpler theories of special materials is more difficult than might at first appear. The additional conditions imposed upon the stress functions by particular constitutive equations usually turn out to be so complicated as to baffle attempts to satisfy them. Consider, for example, the condition of plane isochoric irrotational flow subject to hydrostatic pressure. All fields of velocity and pressure of this class are included among those given by (7.1). The simple conditions $\dot{x}^{\gamma}_{,\gamma}=0$ and $\dot{x}_{[\gamma,\delta]}=0$ when applied to (7.1) yield a complicated non-linear system of equations to be satisfied by the stress functions $a_{\delta\varphi}$ and $a_{3\varphi}$; the classical formulation in terms of velocity-potential and stream-function is preferable in every way. In general, the *more* special conditions are imposed, the less practical is a solution based upon use of stress functions of the second kind.

A more promising application, initiated by Filonenko-Borodich[26] for the case of a rectangular parallelepiped, consists in determining the most general state of stress consistent with assigned values of the stress vector upon a certain boundary. Such results can lead to valuable comparisons, estimates, and mean values by which the results of different theories of materials, or of all possible theories, are compared. *Cf.* the remarks at the end of § 1.

Part II. Curved spaces

10. Formal intrinsic solution in an affine space. In a flat space, the similarity in form between the conditions of compatibility and the general solutions (5.4) and (5.5), or (5.3), has been remarked for half a century[27]. This similarity may be used to yield a direct solution[28] of (2.1) based upon the classical principle of virtual work. Indeed, we shall need only a special case. If $T^{km}=T^{mk}$, for any vector c_k in an n-dimensional space we have

$$\int_V T^{km}\, c_{(k,m)}\, dV = \oint_S T^{km}\, c_k\, da_m - \int_V T^{km}{}_{,m}\, c_k\, dV. \tag{10.1}$$

[26] [1951, 1 & 2] [1957, 2]. The work rests on trigonometric series or special functional forms.

[27] Specifically, the tensor stress function is indeterminant to within a tensor satisfying the conditions of compatibility. A similar observation formed the basis from which B. Finzi [1934, 2, §§ 4, 7] conjectured his stress functions for spaces of constant curvature. See § 12, below. It follows from the results we are about to present, however, that this method of conjecture is valid only when the compatibility operator is self-adjoint, as is the case in flat spaces and in spaces of constant curvature.

[28] The method was given by Truesdell [1957, 5] for the two-dimensional case, revising work of L. Finzi [1956, 3]. Earlier Schaefer [1953, 5, § 4] had introduced multipliers in just the same way, concluding that „Jeder Verträglichkeitsbedingung ist eine Spannungsfunktion zugeordnet," but his presentation employs results of a kind valid only in flat spaces, and he did not mention any further possibilities. Günther [1954, 1, § 2] in effect noted the method and remarked that the tensor of stress functions in a three-dimensional flat space may be interpreted as Lagrangian multipliers expressing the reactions against the geometrical constraints but concluded that "no new point of view results".

Successful treatment of the converse problem is older. As was observed by Locatelli [1940, 1 & 2] for a flat space and more generally by L. Finzi [1956, 3, § 10], if the general solution of the equations of equilibrium in terms of stress functions is known, the form of the conditions of compatibility may be derived from the converse of the principle of virtual work, which in an elastic context is named after Menabrea or Castigliano. It was on this basis that L. Finzi correctly conjectured for membranes the forms of solution presented in § 13, below.

This identity, being based only upon Green's theorem, is valid and invariant in any space where covariant differentiation is defined. For a fixed region V, it follows from (10.1) that *a symmetric tensor* T^{km} *satisfies* $T^{km}{}_{,m}=0$ *in* V *if and only if*

$$\int_V T^{km} d_{km}\, dv = 0 \tag{10.2}$$

for all tensors d_{km} *such that the system*

$$d_{km} = c_{(k,m)} \tag{10.3}$$

has a solution c_k *satisfying the condition* $T^{km} c_k\, da_m = 0$ *upon* S. Since we are interested in satisfying the differential equations everywhere, we may adjust the region so that the boundary condition is satisfied[29]; in effect, we need only consider the interior condition[30] (10.3). In other words, the variational condition (10.2) for all d_{km} such that the system (10.3) is compatible is equivalent to the differential equation $T^{km}{}_{,m}=0$, provided $T^{km}=T^{mk}$.

Let it be supposed that the conditions of compatibility for (10.3), in the space considered, are of the form

$$\begin{aligned} 0 &= \mathscr{L}_{ijkl}(\boldsymbol{d}), \\ &= \alpha_{ijkl}{}^{pq} d_{pq} + \beta_{ijkl}{}^{pqr} d_{pq,r} + \\ &\quad + \gamma_{ijkl}{}^{pqrs} d_{pq,rs} + \delta_{ijkl}{}^{pqrst} d_{pq,rst} + \\ &\quad + \cdots, \end{aligned} \tag{10.4}$$

where the sum is finite. By a standard device, this system of side conditions is taken into account by use of multipliers P^{ijkl}, so that the restricted varational condition (10.2) is replaced by the modified condition

$$\int_V [T^{km} d_{km} - P^{ijkl}\, \mathscr{L}_{ijkl}(\boldsymbol{d})]\, dv = 0 \tag{10.5}$$

for unrestricted $\boldsymbol{d}$. By applying Green's theorem, we put (10.5) into the form

$$\int_V [T^{km} - \bar{\mathscr{L}}^{km}(\boldsymbol{P})]\, d_{km}\, dv + \oint_S \cdots, \tag{10.6}$$

where the dots stand for a surface integral not affecting the result we seek, and where $\bar{\mathscr{L}}$ is the operator adjoint to $\mathscr{L}$:

$$\bar{\mathscr{L}}^{km}(\boldsymbol{P}) = \alpha_{rstu}{}^{km} P^{rstu} - (\beta_{rstu}{}^{kmq} P^{rstu})_{,q} + (\gamma_{rstu}{}^{kmqv} P^{rstu})_{,qv} - \cdots. \tag{10.7}$$

Since (10.6) is to hold for arbitrary d_{km}, it follows that

$$T^{km} = \bar{\mathscr{L}}^{km}(\boldsymbol{P}). \tag{10.8}$$

[29] *Cf.* also the argument used by DORN & SCHILD [1956, 2].

[30] It is worth noticing that from (10.1) there follows another characterization which is not variational, as follows: For a given symmetric T^{km} such that $T^{km}{}_{,m}$ is continuous, let the system (10.3) be such that in any neighborhood of P there exists a region V with boundary S such that (10.3) admits n linearly independent solutions c_k which vanish upon S. Then $T^{km}{}_{,m}=0$ at P if and only if (10.2) holds for all such V.

The condition of symmetry $(2.1)_2$ has not been considered explicitly, but it is obviously satisfied because the tensors $\boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots$ occurring in (10.4) may be taken as symmetric in p and q without loss of generality, so that $\mathscr{L}^{km} = \mathscr{L}^{mk}$. Thus the result (10.8) always yields a symmetric $\boldsymbol{T}$. Conversely, the necessity of (10.8) follows from (10.2), and this in turn is equivalent to $(2.1)_2$ only if $\boldsymbol{T}$ is assumed to be symmetric.

Thus our result (10.8) is the *general solution of the equations of motion* (2.1) in an affine n-dimensional space. The problem of solving (2.1) is thereby reduced to the problem of finding the explicit form (10.4) for the conditions of compatibility of the differential system (10.3), in the particular space considered.

When we attempt to phrase the result as a rigorous theorem subject to precise conditions of regularity, we run up against an unexpected difficulty. While the use of multipliers to eliminate constraints which are partial differential equations goes back to LAGRANGE, is standard practice in continuum mechanics, and has been used without comment by at least one eminent pure mathematician[31], the modern literature seems to be absolutely silent regarding this subject. A straightforward attack reduces the problem to one of existence of solution of (10.8), regarded as a partial differential equation satisfied by $\boldsymbol{P}$ when $\boldsymbol{T}$ is given. However, there are indications that the entire approach through the principle of virtual work ought properly to be regarded in terms of a principle of invariance. Such a principle will not be sought here; for the present, we recognize that the solution (10.8) is to be regarded as formal.

11. The form of the conditions of compatibility in a Riemannian space. In the previous section, the general problem of solving the system (2.1), in an affine space, has been reduced to that of calculating explicitly the conditions of compatibility for the system (10.3) in the same space. For completeness, we indicate the nature of this problem of differential elimination and summarize the general information concerning it available at present.

Our problem is to eliminate c_k between (10.3) and its first p derivatives, for some finite p, where

$$c_{k,m} \equiv \frac{\partial c_k}{\partial x^m} - \Gamma^p_{km}\, c_p. \tag{11.1}$$

Regard the field c_k as given, and let d_{km} be defined by (10.3); put $w_{km} \equiv c_{[k,m]}$. For a symmetric connection $\boldsymbol{\Gamma}$ we have the identities

$$\begin{aligned} &w_{km,p} + w_{mp,k} + w_{pk,m} = 0, \\ &c_{k,mp} - c_{k,pm} = c_q R^q{}_{kmp}, \\ &h_{km,ps} - h_{km,sp} = h_{kq} R^q{}_{mps} + h_{qm} R^q{}_{kps}. \end{aligned} \tag{11.2}$$

Since $c_{k,mq} = d_{km,p} + w_{km,p}$, by forming $c_{k,mp} - c_{k,pm}$ and using $(11.2)_{1,2}$ we find that

$$w_{mp,k} = d_{km,p} - d_{kp,m} - R^q{}_{kmp}\, c_q. \tag{11.3}$$

---

[31] E. HELLINGER, §§ 3d and 7e of: Die allgemeinen Ansätze der Mechanik der Kontinua. Enz. Math. Wiss. IV 2.2, H. 5, 601–694 (1913): „so kann man nach dem bekannten Kalkül der Variationsrechnung ... zugehörige Lagrangesche Faktoren ... einführen ...".

Now forming $w_{mp,ks} - w_{mp,sk}$ and using $(11.2)_3$, we obtain the identity[32]

$$\begin{aligned} & d_{km,ps} - d_{kp,ms} - d_{sm,pk} + d_{sp,mk} + \\ & + (R^q{}_{smp,k} - R^q{}_{kmp,s})\, c_q + R^q{}_{smp}\, d_{qk} - R^q{}_{kmp}\, d_{qs} - \\ & - R^q{}_{kmp}\, w_{qs} + R^q{}_{smp}\, w_{qk} + R^q{}_{pks}\, w_{qm} - R^q{}_{mks}\, w_{qp} = 0. \end{aligned} \tag{11.4}$$

For a flat space, (11.4) reduces to the Kirchhoff-St. Venant conditions

$$d_{km,ps} - d_{kp,ms} - d_{sm,pk} + d_{sp,mk} = 0. \tag{11.5}$$

In more general spaces, we are to derive from (11.4) and from the geometry of the space further identities, by means of which c_q and w_{km} are to be eliminated.

PALATINI[33] has indicated how the calculation can be initiated in the case of a metric space. First, we replace sums of the type $R^q \ldots a_q$ by corresponding sums $R_q \ldots a^q$, then use the relation $R_{kmps} = R_{pskm}$ and the Bianchi identities so as to obtain $(R^q{}_{smp,k} - R^q{}_{kmp,s})\, c_q = -R_{mpsk,q}\, c^q$. Thus (11.4) becomes

$$\begin{aligned} & \delta^{rq}_{kp}\, \delta^{tu}_{ms}\, d_{rt,qu} - R_{mpsk,q}\, c^q - R_{mpsq}\, d^q_k + R_{mpkq}\, d^q_s + \\ & + R_{mpkq}\, w^q{}_s - R_{mpsq}\, w^q{}_k - R_{kspq}\, w^q{}_m + R_{ksmq}\, w^q{}_p = 0. \end{aligned} \tag{11.6}$$

Raising the index k and then contracting upon k and m yields[34]

$$I_{,ps} + d^{\ k}_{ps,\ ,k} - 2d^q{}_{(p,s)q} - 2d^q{}_{(p}\, R_{s)q} - 2w^q{}_{(p}\, R_{s)q} - R_{ps,q}\, c^q = 0, \tag{11.7}$$

where $(11.2)_3$ has been used, where $R_{pq} \equiv R^h{}_{pqh}$, and where $I \equiv d^k_k$. Raising the index p in (84.10) and then contracting upon q and s yields[35]

$$I^{,p}_{,p}\, d^{kp}{}_{,kp} - R_{pq}\, d^{pq} = R_{,q}\, c^q \tag{11.8}$$

where $R \equiv R^k_k$.

Equations (11.6) are to be regarded as a system of $\frac{1}{2}n(n+1)$ non-homogeneous linear equations for the $\frac{1}{2}n(n+1)$ unknowns c^q and w^{km} in terms of the tensor $\boldsymbol{d}$ and its second derivatives. Substituting the solution of these equations into (11.6) yields, in principle, a set of differential equations for $\boldsymbol{d}$ alone[36]. The geometrical meaning of the condition of solubility expressed in terms of a determinant formed from the components R_{pq} and $R_{pq,s}$ is not known; of course, it is not satisfied in a flat space, nor is it relevant, since the appropriate condition has already been shown to be (11.5).

More generally, directly from (10.3) we see that the nature of the solution depends very strongly upon the geometry of the space considered. In spaces

[32] RICCI [1888, Eq. (20)].

[33] [1934, 3]. The method was sufficiently indicated by his earlier analysis of a space of constant curvature [1916]. *Cf.* also ANDRUETTO [1932, 1 and 2], AGOSTINELLI [1933], GRAIFF [1958, 1, §§ 6–9].

[34] This identity was given in a special co-ordinate system by PALATINI [1934, 3]; an equivalent general form is due to GRAIFF [1958, 1, Eq. (22)].

[35] GRAIFF [1958, 1, Eq. (23)].

[36] This seems to be a correct substitute for the method of PALATINI [1934, 3], who notes that it is possible to choose co-ordinates so that at a fixed point (11.7) with $p = s$ becomes a system of n linear equations for the n components c^q but fails to note that the solution of this system is not sufficient, in general, to obtain the derivatives of c^q and so to determine w^{km}.

admitting a group of "motions", *i.e.*, instantaneously rigid velocity fields, the solution of (10.3), if compatible, is determinate only to within such a motion. In spaces where instantaneously rigid motions are not possible, the system (10.3) will determine $\boldsymbol{c}$ uniquely from a given $\boldsymbol{d}$, if compatible. Signorina GRAIFF[37] has initiated a discussion of the problem in terms of the principal invariants of the tensor R_{pq}, according as these are linearly independent and non-constant or not. The details are complex, and the problem must still be regarded as open.

In a space of constant curvature, we have

$$\begin{aligned} n(n-1)\,R^q{}_{kmp} &= R(g_{km}\,\delta^q_p - g_{kp}\,\delta^q_m) \\ n\,R_{km} &= R_{qkm}, \end{aligned} \tag{11.9}$$

where $R \equiv K^k_k = \text{const}$. The terms involving $\boldsymbol{c}$ and $\boldsymbol{w}$ in (11.6) vanish. The resulting conditions, easily seen to be sufficient as well as necessary, are[38]

$$\delta^{rq}_{kp}\,\delta^{tu}_{ms}\left(d_{rt,qu} - \frac{R}{n(n-1)}\,d_{rt}\,g_{qu}\right) = 0. \tag{11.10}$$

When $n=2$, there is but one linearly independent non-vanishing component of (11.10), which may be written in the equivalent forms

$$\begin{aligned} e^{\alpha\beta}\,e^{\gamma\delta}\,d_{\alpha\gamma,\beta\delta} + K\,I &= 0, \\ I^{,\alpha}_{,\alpha} - d^{\alpha\beta}{}_{,\alpha\beta} + K\,I &= 0, \end{aligned} \tag{11.11}$$

where K is the total curvature, $K = -\frac{1}{2}R$.

More generally, when $n=2$ and K is arbitrary we have $R_{\alpha\beta\gamma\delta} = K e_{\alpha\beta}\,e_{\gamma\delta}$ and $R_{\alpha\beta} = -K a_{\alpha\beta}$, where $\boldsymbol{a}$ is the metric tensor. Again the terms containing $\boldsymbol{w}$ in (11.6) are annulled, and we have

$$e^{\alpha\beta}\,e^{\gamma\delta}\,d_{\alpha\gamma,\beta\delta} + K\,I = -2K_{,\alpha}\,c^{\alpha}. \tag{11.12}$$

Further differentiations are needed to eliminate $\boldsymbol{c}$. B. FINZI[39] has shown that for a surface applicable upon a surface of revolution the final condition assumes the form

$$(K_{,\gamma}\,K^{,\gamma}\,a^{\alpha\beta} + K^{,\alpha}\,K^{,\beta})\,d_{\alpha\beta} + K\,K^{,\gamma}\,d^{\alpha\beta}\,d_{\alpha\beta,\gamma} + e^{\alpha\beta}\,e^{\gamma\delta}\,K^{,\varepsilon}\,d_{\beta\delta,\alpha\gamma\varepsilon} = 0. \tag{11.13}$$

For a general surface, he obtained a system of three conditions of compatibility, each being a differential equation of fourth order. I remarked[40] that they must be equivalent to a single condition of fifth order. By a method of infinitesimal variation applied to a complete set of differential invariants of the surface in question, Signorina GRAIFF[41] has given a definitive treatment of the problem. She has found two identities relating the quantities occurring in B. FINZI's conditions, but she has not effected the elimination explicitly. She has charac-

[37] [1958, 1, § 10].

[38] PALATINI [1916], [1934, 3], ANDRUETTO [1932, 2], FINZI [1934, 2] including a simplification when $n=3$.

[39] [1930, 1].

[40] [1957, 6]; the reasoning is presented in § 13, below.

[41] [1957, 3].

terized the case when the least possible order of the single scalar condition is 4; in this case, as for a surface of revolution, the lines $K=\text{const.}$ are geodesic parallels, but $K^{\alpha}_{;\alpha}$ is not a function of K only.

Even in the cases when $n=3$ and $n=4$, the possibilities are various[42], and the details are too intricate to summarize here. The fullest information now available is included in the two papers of Signorina GRAIFF already cited.

12. Application to spaces of constant curvature. In a space of constant curvature, from (11.10) it follows that the coefficients in the form (10.4) have the special forms

$$\begin{aligned} \alpha_{ijkl}{}^{pq} &= -\frac{R}{n(n-1)}\,\delta^{pu}_{ij}\,\delta^{qv}_{kl}\,g_{uv}, \\ \gamma_{ijkl}{}^{pqrs} &= \delta^{pr}_{ij}\,\delta^{qs}_{kl}, \end{aligned} \tag{12.1}$$

while all other coefficients vanish. The operator $\mathscr{L}$ is self-adjoint, and from (10.8) we have the following *general solution for a space of constant curvature:*

$$T^{km} = h^{pkmq}{}_{,pq} - \frac{R}{n(n-1)}\,h^{qkm}{}_{q}, \tag{12.2}$$

where we have set

$$h^{pkmq} \equiv \delta^{kp}_{rs}\,\delta^{mq}_{tu}\,P^{rstu}, \tag{12.3}$$

$\boldsymbol{P}$ being the tensor in terms of which the general solution of § 10 is expressed. From (12.3), it is immediately plain that $\boldsymbol{h}$ satisfies the conditions of symmetry $(4.7)_{2,3,4}$. Thus the solution (4.6) for flat spaces is included in (12.2) as the special case when $R=0$.

When $n=3$, we may introduce the tensor $\boldsymbol{a}$ dual to $\boldsymbol{h}$ according to (5.1). The solution (12.2) now becomes[43]

$$\begin{aligned} T^{km} &= e^{krp}\,e^{msq}\left(a_{rs,pq} - \tfrac{1}{6}\,R\,a_{rs}\,g_{pq}\right), \\ &= e^{krp}\,e^{msq}\,a_{rs,pq} - \tfrac{1}{6}\,R\left(g^{km}\,a^{q}_{q} - a^{km}\right), \end{aligned} \tag{12.4}$$

generalizing (5.3).

When $n=4$, we may introduce the tensor $\boldsymbol{A}$ dual to $\boldsymbol{h}$ according to (8.1) and so obtain from (12.2) the solution[44]

$$T^{\Gamma\Delta} = e^{\Gamma\Lambda\Sigma\Phi}\,e^{\Delta\Xi\Psi\Omega}\left(A_{\Sigma\Phi\Psi\Omega,\Xi\Lambda} - \tfrac{1}{12}\,R\,A_{\Sigma\Phi\Psi\Omega}\,g_{\Xi\Lambda}\right), \tag{12.5}$$

generalizing (8.3). However, this form of the solution does not seem to offer any advantage over (12.2).

13. Application to the classical theory of membranes. We now consider the membrane problem: To find the general solution of (3.10), where covariant differentiation is based upon a given two-dimensional Riemannian metric $\boldsymbol{a}$,

---

[42] GRAIFF [1958, 1, §§ 13—14].

[43] B. FINZI [1934, 2, § 4].

[44] B. FINZI [1934, 2, § 7] gives the solution

$$T^{\Gamma\Delta} = e^{\Gamma\Lambda\Sigma\Phi}\,e^{\Delta\Xi\Psi\Omega}\,A_{\Sigma\Phi\Psi\Omega,\Lambda\Xi} + K(A_{\Sigma\Phi}{}^{\Sigma\Phi}\,g^{\Gamma\Delta} - 2A^{\Gamma}{}_{\Phi}{}^{\Delta\Phi});$$

this may be equivalent to (12.5) but is not obviously so.

and where we take $F^\delta = 0$. By direct and laborious reduction, STORCHI[45] has obtained a general solution in geodesic co-ordinates. This solution involves derivatives of orders up to 5 of a single stress function. At the present writing, a corresponding invariant solution is not yet known. In order to find it by our method, we should have to perform explicitly the reduction of B. FINZI's conditions of compatibility to a single scalar equation.

We rest content with recording the simple cases that follow effortlessly from the results already given. When the condition of compatibility is a single scalar condition $\mathscr{L}(\boldsymbol{d}) = 0$, the analysis proceeds just as in § 10, except that the four indices $i\,j\,k\,l$ are suppressed, and the multiplier reduces to a single scalar A.

For a surface of constant curvature, by comparing (11.10) with the reduced form of (10.4) we see that

$$\alpha^{\sigma\varphi} = K a^{\sigma\varphi}, \quad \beta^{\sigma\varphi\varrho} = 0, \quad \gamma^{\sigma\varphi\varrho\tau} = e^{\sigma\varrho} e^{\varphi\tau}, \tag{13.1}$$

and all other coefficients vanish. Substitution into the reduced form of (10.8) yields the *general solution for surfaces of constant curvature*[46]:

$$T^{\alpha\beta} = \epsilon^{\alpha\gamma} \epsilon^{\beta\varphi} A_{,\gamma\varphi} + K A\, a^{\alpha\beta}. \tag{13.2}$$

Of course, this same result follows from (12.2) and the observation that when $n = 2$, the most general tensor h having the symmetries $(4.7)_{2,3,4}$ is of the form $h^{\gamma\alpha\beta\delta} = e^{\alpha\gamma} e^{\beta\delta} A$ for some scalar A.

For a surface applicable upon a surface of revolution, by comparing (11.12) with the reduced form of (10.4) we see that

$$\begin{aligned} &\alpha^{\varrho\varphi} = K_{,\sigma} K^{,\sigma} a^{\varrho\varphi} + K^{,\varrho} K^{,\varphi}, \\ &\beta^{\varphi\varrho\sigma} = K K^{,\sigma} a^{\varrho\varphi}, \quad \gamma^{\varrho\varphi\sigma\tau} = 0, \\ &\delta^{\varrho\varphi\sigma\tau\nu} = -e^{\tau(\varrho} e^{\varphi)\sigma} K^{,\nu}, \end{aligned} \tag{13.3}$$

and all other coefficients vanish. Substitution into the reduced form of (10.8) yields the *general solution for a surface applicable upon a surface of revolution*

[45] [1950, 5]. STORCHI [1950, 4] observed also that if for an arbitrary surface we choose co-ordinates x, y so that $ds^2 = \lambda(dx^2 + dy^2)$, then a solution of (2.10) when $\boldsymbol{F} = 0$ is furnished by the formulae $\lambda^2 S^{xx} = \epsilon^2 A/\partial y^2$, $\lambda^2 S^{xy} = -\epsilon^2 A/\partial x\, \partial y$, $\lambda^2 S^{yy} = \partial^2 A/\partial x^2$, provided that $\epsilon^2 A/\partial x^2 + \epsilon^2 A/\partial y^2 = 0$. For such solutions the mean pressure vanishes: $\lambda(S^{xx} + S^{yy}) = S^\delta_\delta = 0$, but it is not shown that all solutions such that $S^\delta_\delta = 0$ are included. STORCHI treated the case of a surface of revolution in [1949, 3]; in [1952, 2] and [1953, 6], the case when a general solution involving derivatives of orders no higher than the fourth is possible. *Cf.* the corresponding condition of compatibility mentioned at the end of § 11. A solution for minimal surfaces is initiated by COLONNETTI [1956, 1]. I do not attempt to present the solutions, some classical and some recent, for special cases defined by conditions of inextensibility. For Riemannian spaces of 2, 3, and 4 dimensions, a special solution containing an arbitrary function of the total curvature is obtained by STORCHI [1957, 4].

[46] B. FINZI [1934, 2, § 2] verified that (13.2) satisfies (2.1) when K is constant but did not prove the completeness of this solution; his result is rediscovered by LANGHAAR [1953, 2].

of non-constant curvature[47]:

$$\begin{aligned} T^{\alpha\beta} - [&- K K_{,\gamma}^{,\gamma} a^{\alpha\beta} + K^{,\alpha} K^{,\beta} + e^{\delta(\alpha} \epsilon^{\beta)\gamma} K^{,\varepsilon}{}_{,\gamma\delta\varepsilon}] A + \\ &+ [- K K^{,\gamma} a^{\alpha\beta} + e^{\varepsilon(\alpha} e^{\beta)\delta} K^{,\gamma}{}_{,\delta\varepsilon} + 2 e^{\gamma(\alpha} e^{\beta)\delta} K^{,\varepsilon}{}_{,\delta\varepsilon}] A_{,\gamma} + \\ &+ [2 e^{\gamma(\alpha} e^{\beta)\delta} K^{,\varepsilon}{}_{,\delta} + e^{\varepsilon(\alpha} e^{\beta)\gamma} K_{,\lambda}^{,\lambda}] A_{,\gamma\varepsilon} + \\ &+ e^{\delta(\alpha} \epsilon^{\beta)\gamma} K^{,\varepsilon} A_{,\gamma\delta\varepsilon}. \end{aligned} \tag{13.4}$$

[*Note added in proof, September 15, 1959.* The elegant paper of GRAIFF [1959, 3] on this problem has reached me too late for a full description to be included here. Her essential idea is to begin with an invariant decomposition of the three-dimensional space of symmetric covariant tensors defined upon the surface. The general case is defined as that in which the scalars K and $H \equiv K_{,\alpha} K^{,\alpha}$ are functionally independent. Then the three symmetric covariant tensors $K_{,\alpha} K_{,\beta}$, $H_{,\alpha} H_{,\beta}$, and $H_{(,\alpha} K_{,\beta)}$ form a basis for the symmetric covariant tensors $F_{\alpha\beta}$:

$$F_{\alpha\beta} = A K_{,\alpha} K_{,\beta} + B H_{(,\alpha} K_{,\beta)} + C H_{,\alpha} H_{,\beta}, \tag{13.5}$$

where the scalar functions A, B, C are uniquely determined joint invariants of F, K, and H which are easy to exhibit. After calculation of $F_{\alpha\beta}{}^{,\beta}$ it turns out that the most general symmetric tensor $F_{\alpha\beta}$ such that

$$F_{\alpha\beta}{}^{,\beta} \propto K_{,\alpha} \tag{13.6}$$

corresponds to a special choice of A in (13.5). Now let φ be any scalar function, and put

$$\Phi_{\alpha\beta} = a_{\alpha\beta}(\varphi_{,\gamma}{}^{,\gamma} + K\varphi) - \varphi_{,\alpha\beta}; \tag{13.7}$$

then

$$\Phi_{\alpha\beta}{}^{,\beta} = \varphi K_{,\alpha}. \tag{13.8}$$

Hence the most general solution of (2.1) is of the form $T_{\alpha\beta} = D_{\alpha\beta} + \Phi_{\alpha\beta}$ where $D_{\alpha\beta}$ is the most general solution of

$$D_{\alpha\beta}{}^{,\beta} = -\varphi K_{,\alpha} \tag{13.9}$$

for any given φ. Matching (13.6) and (13.9) determines a unique value for φ. The result is an explicit formula for the general solution of (2.1) in terms of derivatives *up to fourth order* of *two scalar potentials*. This is a result of the type to be expected in view of the conditions of compatibility discussed in § 11.

However, GRAIFF infers that *either one of the two scalars may always be taken as zero*. This contradicts the results of STORCHI, cited in footnote 45, according to which it is only an exceptional class of surfaces for which fourth derivatives of a single scalar suffice. In GRAIFF's work, a solution χ to her equation $(23)_1$ with $B' = 0$ or $(23)_2$ with $C' = 0$, for any B and C, respectively, is assumed, but the existence of solution to this problem on a curved surface is not proved.]

An essentially different approach to the membrane problem has been initiated by PUCHER[48]. He combines (3.10) and (3.11) from the start, so that the intrinsic problem is never solved, and he uses special oblique co-ordinates on the surface. While a simpler derivation of his results has been given by L. FINZI[49], it does not seem easy to put the analysis into invariant form.

14. Status of the problem for curved spaces. While the powerful method of § 10 reduces the problem to one of differential elimination, the necessary

[47] Given in geodesic co-ordinates, with a proof of completeness, by STORCHI [1949, 3]; in the above form, with an imperfect proof of completeness, by L. FINZI [1956, 3].

[48] [1934, 5] [1939].

[49] [1955, 1].

calculations seem to be difficult. For physical interpretation, curved spaces are most interesting in two special cases: ordinary surfaces and four-dimensional space-time.

The most useful curved surfaces are included in the results given in § 13; the one further elimination needed for the general surface is known to be possible and doubtless will soon be effected.

No special simplification is apparent, however, when $n=4$. Furthermore, recent studies of the foundations of mechanics show that the most interesting result would be that for a not necessarily Riemannian affine space. While RICCI's identity (11.4) is valid for such a space, none of the further calculations in § 11 are applicable. Since the problem as stated is an affine one, it seems that the entry so far used is not really cogent, and that a more penetrating study of the structure of affine spaces is needed.

Acknowledgment. I am grateful to Professor ERICKSEN for assistance in preparing § 8 and to Professor STERNBERG for criticism of an earlier version.

This report was written for the Rheology Section, National Bureau of Standards, Washington, D.C.

References

1757 EULER, L.: Continuation des recherches sur la théorie du mouvement des fluides. Mém. acad. sci. Berlin [**11**], 316—361 (1755) = Opera omnia (2) **12**, 92—132.

1863 AIRY, G. B.: On the strains in the interior of beams. Phil. trans. r. soc. London **153**, 49—80. Abstract in Rep. Brit. Assn. **1862**, 82—86 (1863).

1868 MAXWELL, J. C.: On reciprocal diagrams in space, and their relation to AIRY's function of stress. Proc. London Math. Soc. (1) **2**, 58—60 (1865—1869) = Papers **2**, 102—104.

1870 MAXWELL, J. C.: On reciprocal figures, frames, and diagrams of forces. Trans. r. soc. Edinburgh **26**, 1—40 (1869—1872) = Papers **2**, 161—207.

1883 VOIGT, W.: Allgemeine Formeln für die Bestimmung der Elasticitätsconstanten von Krystallen durch die Beobachtung der Biegung and Drillung von Prismen. Annalen der Phys. (2) **16**, 273—321, 398—416.

1888 RICCI, G.: Delle derivazioni covarianti e controvarianti e del loro uso nella analisi applicata. Studi univ. Padova comm. ottavo centenar. Univ. Bologna **3**, No. 12, 23 pp.

1892 1. MORERA, G.: Soluzione generale delle equazioni indefinite dell'equilibrio di un corpo continuo. Rend. Lincei (5) $\mathbf{1}_1$, 137—141.

2. BELTRAMI, E.: Osservazioni sulla nota precedente. Rend. Lincei (5) $\mathbf{1}_1$, 141—142 = Opere **4**, 510—512.

3. MORERA, G.: Appendice alla Nota: sulla soluzione più generale delle equazioni indefinite dell'equilibrio di un corpi continuo. Rend. Lincei (5) $\mathbf{1}_1$, 233—234.

1900 1. MICHELL, J. H.: On the determination of stress in an elastic solid, with applications to the theory of plates. Proc. London Math. Soc. **31**, 100—124 (1899).

2. MICHELL, J. H.: The uniform torsion and flexure of incomplete tores, with application to helical springs. Proc. London Math. Soc. **31**, 130—146 (1899).

1905 KLEIN, F., & K. WIEGHARDT: Über Spannungsflächen und reziproke Diagramme, mit besonderer Berücksichtigung der Maxwellschen Arbeiten. Arch. Math. Phys. (3) **8**, 1—10, 95—119.

1906 LOVE, A. E. H.: A Treatise on the Mathematical Theory of Elasticity, 2nd ed. Cambridge.

1907 NEUMANN, E. R.: Über eine neue Reduktion bei hydrodynamischen Problemen. J. reine angew. Math. **132**, 189—215.

1911 GWYTHER, R. R.: The conditions that the stresses in a heavy elastic body should be purely elastic stresses. Mem. Manchester lit. phil. soc. **55**, No. 20 (12 pp.).

1912 GWYTHER, R. R.: The formal specification of the elements of stress in Cartesian, and in cylindrical and spherical polar coordinates. Mem. Manchester lit. phil. soc. **56**, No. 10 (13 pp.).

1913 GWYTHER, R. F.: The specification of the elements of stress, Part II. A simplification of the specification given in Part I. Mem. Manchester lit. phil. soc. **57**, No. 5 (4 pp.).

1916 PALATINI, A.: Sulle quadriche di deformazione per gli spazi S_3. Atti ist. Veneto **76**, 125—148.

1930 1. FINZI, B.: Sopra il tensore di deformazione di un velo. Ist. Lombardo rend. (2) **63**, 975—982.

2. KIRCHHOFF, W.: Reduktion simultaner partiellen Differentialgleichungen bei hydrodynamischen Problemen. J. reine angew. Math. **164**, 183—195.

1932 1. ANDRUETTO, G.: Le formule di SAINT-VENANT per gli spazi curvi a tre dimensioni. Rend. accad. Lincei (6) **15**, 214—218.

2. ANDRUETTO, G.: Le formule di SAINT-VENANT per le varietà V_n a curvatura costante. Rend. accad. Lincei (6) **15**, 792—797.

1933 AGOSTINELLI, C.: Le condizioni di SAINT-VENANT per le deformazioni di una varietà riemanniana generica. Rend. accad. Lincei (6) **18**, 529—533; (6) **19**, 22—26 (1934).

1934 1. BRAHTZ, J.: Notes on the AIRY stress function. Bull. Amer. math. soc. **40**, 427—430.

2. FINZI, B.: Integrazione delle equazioni indefinite della meccanica dei sistemi continui. Rend. accad. Lincei (6) **19**, 578—584, 620—623.

3. PALATINI, A.: Sulle condizioni di SAINT-VENANT in una V_n qualsivoglia. Rend. accad. Lincei (6) **19**, 466—469.

4. PHILLIPS, H. B.: Stress functions. J. math. phys. M. I. T. **13**, 421—425.

5. PUCHER, A.: Über den Spannungszustand in gekrümmten Flächen. Beton u. Eisen **33**, 298—304.

1935 SOBRERO, L.: Del significato meccanico della funzione de AIRY. Ricerche di ingegn. **3**, 77—80 = Rend. Lincei (6) **21**, 264—269.

1936 BATEMAN, H.: Progressive waves of finite amplitude and some steady motions of an elastic fluid. Proc. nat. acad. sci. U.S.A. **22**, 607—619.

1938 BATEMAN, H.: The lift and drag functions for an elastic fluid in two dimensional irrotational flow. Proc. nat. acad. sci. U.S.A. **24**, 246—251.

1939 PUCHER, A.: Über die Spannungsfunktion beliebig gekrümmter dünner Schalen. Proc. 5th. int. congr. applied mech., Cambridge 1938, 134—139.

1940 1. LOCATELLI, P.: Sulla congruenza delle deformazioni. Rend. ist. Lombardo **13** = (3) **4**, 451—464 (1939—1940).

2. LOCATELLI, P.: Sul principio di MENABREA. Boll. un. mat. Ital. (2) **2**, 342—347.

1945 KUZMIN, R. O.: On MAXWELL's and MORERA's formulae in the theory of elasticity. C. r. (Doklady) acad. sci. URSS **49**, 326—328.

1948 WEBER, C.: Spannungsfunktionen des dreidimensionalen Kontinuums. Z. angew. Math. Mech. **28**, 193—197.

1949 1. FINZI, B., & M. PASTORI: Calcolo tensoriale e applicazioni. Bologna, Zanichelli.

2. PERETTI, G.: Significato del tensore arbitrario che interviene nell' integrale generale delle equazioni della statica dei continui. Atti sem. mat. fis. univ. Modena **3**, 77—82.

3. STORCHI, E.: Integrazione delle equazioni indefinite della statica dei sistemi continui su una superficie di rotazione. Rend. accad. Lincei (7) **7**, 227—231.

1950 1. BLOKH, V.: Stress functions in the theory of elasticity [in Russian]. Prikl. Mat. Mekh. **14**, 415—422.

2. CROCCO, L.: On a kind of stress-function for the study of non-isentropic two-dimensional motion of gases. Proc. 7th int. congr. appl. mech. London 1948. 2, 315—329.

3. Morinaga, K., & T. Nôno: On stress-functions in general coordinates. J. sci. Hiroshima Univ. A **14**, 181—194.
4. Storchi, E.: Sulle equazioni indefinite della statica delle membrane tese su generiche superficie. Rend. accad. Lincei (7) **8**, 116—120.
5. Storchi, E.: Integrazione delle equazioni indefinite della statica dei veli tesi su una generica superficie. Rend. accad. Lincei (7) **8**, 326—331.

1951 1. Filonenko-Borodich, M. M.: The problem of the equilibrium of an elastic parallelepiped subject to assigned loads on its boundaries [in Russian]. Prikl. Mat. Mekh. **15**, 137—148.
2. Filonenko-Borodich, M. M.: Two problems on the equilibrium of an elastic parallelepiped [in Russian]. Prikl. Mat. Mekh. **15**, 563—574.

1952 1. Arzhanikh, I. C.: Tensor functions of hydrodynamical stresses [in Russian]. Doklad. Akad. Nauk SSSR **83**, 195—198.
2. Storchi, E.: Le superficie eccezionali nella statica delle membrane. Revista mat. univ. Parma **3**, 339—360.

1953 1. Kilchevski, N. A.: Stress, velocity, and density functions in static and dynamic problems in the mechanics of continuous media [in Russian]. Doklad. Akad. Nauk SSSR **92**, 895—898.
2. Langhaar, H.: An invariant membrane stress function for shells. J. appl. mech. **20**, 178—182.
3. McVittie, G. C.: A method of solution of the equations of classical gas dynamics using Einstein's equations. Q. appl. math. **11**, 327—336.
4. Pratelli, A.: Sulla stazionarietà di significitavi integrali nella meccanica dei continui. Rend. ist. Lombardo (3) **17** (86), 714—724.
5. Schaefer, H.: Die Spannungsfunktionen des dreidimensionalen Kontinuums und des elastischen Körpers. Z. angew. Math. Mech. **33**, 356—362.
6. Storchi, E.: Sulle membrane aventi comportamento meccanico eccezionale. Ist. Lombardo rend. (3) **17** (86), 462—483.

1954 1. Günther, W.: Spannungsfunktion und Verträglichkeitsbedingungen der Kontinuumsmechanik. Abh. Braunschweig. Wiss. Ges. **7**, 107—112.
2. Langhar, H., & M. Stippes: Three-dimensional stress functions. J. Franklin inst. **258**, 371—382.
3. Ornstein, W.: Stress functions of Maxwell and Morera. Q. appl. math. **2**, 198—201.
4. Whitham, G. B.: A note on a paper by G. C. McVittie. Q. appl. math. **12**, 316—318.

1955 1. Finzi, L.: Sulle equazioni di Pucher nell'equilibrio delle strutture a guscio. Rend. ist. Lombardo (3) **88**, 907—916.
2. Marguerre, K.: Ansätze zur Lösung der Grundgleichungen der Elasticitätstheorie. Z. angew. Math. Mech. **35**, 242—263.
3. Schaefer, H.: Die Spannungsfunktion einer Dyname. Abh. Braunschweig. Wiss. Ges. **7**, 107—112.
4. Schaefer, H.: Die drei Spannungsfunktionen des zweidimensionalen ebenen Kontinuums. Österr. Ing.-Arch. **10**, 267—277. [I am unable to see the connection found by Prager, Math. Rev. **18**, 613 (1957), between this work and one by Fox & Southwell.]

1956 1. Colonnetti, G.: L'équilibre des voiles minces hyperstatiques (Le cas des voiles de surface minimum). C. r. acad. sci. Paris **243**, 1087—1089, 1701—1704.
2. Dorn, W. S., & A. Schild: A converse to the virtual work theorem for deformable solids. Q. appl. math. **14**, 209—213.
3. Finzi, L.: Legame fra equilibrio e congruenza e suo significato fisico. Rend. accad. Lincei (8) **20**, 205—211, 338—342.

1957 1. Brdička, M.: On the general form of the Beltrami equation and Papkovich's solution of the axially symmetric problem of the classical theory of elasticity. Czechosl. J. phys. **7**, 262—274.
2. Filonenko-Borodich, M. M.: On the problem of Lamé for the parallelepiped in the general case of surface loads [in Russian]. Prikl. Mat. Mekh. **21**, 550—559.

3. Graiff, F.: Sulle condizioni di congruenza per una membrana. Rend. ist. Lombardo (A) **92**, 33—42.
4. Milne-Thomson, L. M.: A general solution of the equations of hydrodynamics. J. Fluid. Mech. **2**, 88.
5. Storchi, E.: Una soluzione delle equazioni indefinite della meccanica dei continui negli spazi riemanniani. Rend. ist. Lombardo **90**, 369—378 (1956).
6. Truesdell, C.: General solution for the stresses in a curved membrane. Proc. nat. acad. sci. U.S.A. **43**, 1070—1072.

1958 1. Graiff, F.: Sulle condizioni di congruenza per deformazioni anche finite. Rend. accad. Lincei (8) **24**, 415—422.
2. Günther, W.: Zur Statik und Kinematik des Cosseratschen Kontinuums. Abh. Braunschw. Wiss. Ges. **10**, 195—213.

1959 1. Blankfield, J., & G. C. McVittie: Einstein's equation and classical hydrodynamics. Arch. Rational Mech. Anal. **2**, 337—354.
2. Blankfield, J., & G. C. McVittie: A method of solution of the equations of magneto-hydrodynamics. Arch. Rational Mech. Anal. **2**, 411—422.
3. Graiff, F.: Soluzione tensoriale generale delle equazioni indefinite di equilibrio per una membrana. Rend. accad. Lincei (8) **26**, 189—196.

Appendix

Bibliography of works on stress functions for linear elasticity and related theories, not repeating items already cited as references to the foregoing article.

1829 E 1. Poisson, S. D.: Addition au mémoire sur l'équilibre et le mouvement des corps élastiques. Mém. Acad. Sci. Paris **8**, 623—627.

1851 E 1. Stokes, G. G.: On the dynamical theory of diffraction. Trans. Cambridge Phil. Soc. **9** (1851—1856), 1—62 = Papers **2**, 243—328.

1852 E 1. Lamé, G.: Leçons sur la Théorie Mathématique de l'Élasticité des Corps Solides. Paris: Bachelier.

1863 E 1. Clebsch, A.: Über die Reflexion an einer Kugelfläche. J. Reine Angew. Math. **61**, 195—262.

1882 E 1. Cerruti, V.: Ricerche intorno all' equilibrio de' corpi elastici isotropi. Atti accad. Lincei Memorie (3) **13**, 81—123.

1885 E 1. Boussinesq, M.: Application des potentiels à l'étude de l'équilibre et des mouvements des solides élastiques. Paris: Gauthier-Villars. See §§ II, VII.

1887 E 1. Ibbetson, W.: On the Airy-Maxwell solution of the equations of equilibrium of an isotropic elastic solid, under conservative forces. Proc. Lond. Math. Soc. **17**, 296—309 (1885—1886).

1889 E 1. Jaerisch, P.: Allgemeine Integration der Elasticitätsgleichungen für die Schwingungen und das Gleichgewicht isotroper Rotationskörper. J. Reine Angew. Math. **104**, 177—210.

E 2. Jaerisch, P.: Zur Theorie der Elasticität isotroper Rotationskörper. Mitteil. Hamburger Math. Ges. **1**, 278—289.

1898 E 1. Duhem, P.: Sur l'intégrale des équations des petits mouvements d'un solide isotrope. Mem. Soc. Sci. Bordeaux (5) **3**, 317—329.

1900 E 1. Duhem, P.: Sur la généralisation d'un théorème de Clebsch. J. Math. Pures Appl. (5) **6**, 215—259.

1904 E 1. Lord Kelvin: Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light (1884). London: C. J. Clay and Sons. See Lecture IV, p. 41 *et sqq.*

E 2. Marcolongo, R.: Teoria Matematica dello Equilibrio dei Corpi Elastici. Milano. See p. 236 *et sqq.*

E 3. Tedone, O.: Saggio di una teoria generale delle equazioni dell' equilibrio elastico per un corpo isotropo. Annali di mat. (3) **10**, 13—64.

1914 E 1. Gwyther, R.: The specification of the elements of stress. Part III. The definition of the dynamical specification and a test of the elastic specification. A chapter in elasticity. Mem. Manchester lit. philos. soc. **58**, No. 5 (21 pp.).

1915 E 1. Korn, A.: Über die beiden bisher zur Lösung der ersten Randwertaufgabe der Elastizitätstheorie eingeschlagenen Wege. Annaes acad. poly. Porto **10**, 1—28.
1918 E 1. Gwyther, R. F.: The specification of stress, Part V. Mem. Manchester lit. phil. soc. **62**, No. 1, 11 pp. (1917—1918).
1919 E 1. Serini, R.: Deformazione simmetriche dei corpi elastici. Atti accad. Lincei (5) **28**, 343—347.
1923 E 1. Burgatti, P.: Sopra una soluzione molto generale dell' equazioni dell' equilibrio elastico. Rend. accad. Bologna (2) **27**, 66—73 (1922—1923).
1924 E 1. Timpe, A.: Achsensymmetrische Deformation von Umdrehungskörpern. Z. angew. Math. Mech. **4**, 361—376.
1925 E 1. Weber, C.: Achsensymmetrische Deformation von Umdrehungskörpern. Z. angew. Math. Mech. **5**, 466—468.
1926 E 1. Burgatti, P.: Sopra due utili forme dell' integrale generale dell' equazioni per l'equilibrio dei solidi elastici isotrope. Mem. accad. Bologna (8) **3**, 63—67.
1928 E 1. Trefftz, E.: Mathematische Elastizitätstheorie. Handbuch der Physik **6**. Berlin: Springer. See p. 91 *et seqq.*
E 2. Levi-Civita, T.: Fondamenti di Meccanica Relastivistica. Bologna: Zanichelli. vii + 185 pp. See §§ 25—26.
1930 E 1. Galerkin, B.: Contribution à la solution générale du problème de la théorie de l'élasticité dans le cas de trois dimensions. C. r. acad. sci. Paris **190**, 1047—1048.
E 2. Galerkin, B.: On an investigation of stresses and deformations in elastic isotropic solids [in Russian]. Doklad. Akad. Nauk SSSR **1930 A**, 353—358.
E 3. Malkin, I.: Über einige neuere Arbeiten auf dem Gebiete der Elastizitätslehre. Z. angew. Math. Mech. **10**, 182—197.
1931 E 1. Galerkin, B.: Sur l'équilibre élastique d'un plaque rectangulaire épaisse. C. r. acad. sci. Paris **193**, 563—571.
E 2. Galerkin, B.: Elastic rectangular and triangular thick plates with free edges, subjected to bending [in Russian]. Doklad. Akad. Nauk SSSR **1931**, 273—280.
E 3. Galerkin, B.: On the general solution of a problem in the theory of elasticity in three dimensions by means of stress and displacement functions [in Russian]. Doklad. Akad. Nauk SSSR 1931, 281—286.
1932 E 1. Galerkin, B.: On the investigation of stresses and deformation in a thick rectangular plate [in Russian]. Isv. Nauchno-Issled. Inst. Gidrotekh. **6**, 28—38.
E 2. Galerkin, B.: General solution of a problem on stresses and deformation in a thick circular plate and in a plate having the form of a circular sector [in Russian]. Isv. Nauchno-Issled. Inst. Gidrotekh. **7**, 1—6.
E 3. Papcovitch, P.: Solution générale des équations différentielles fondamentales d'élasticité, exprimée par trois fonctions harmoniques. C. r. acad. sci. Paris **195**, 513—515.
E 4. Papcovitch, P.: Expressions générales des composantes des tensions, ne renfermant comme fonctions arbitraires que des fonctions harmoniques. C. r. acad. sci. Paris **195**, 754—756.
E 5. Papkovich, P.: The representation of the general integral of the fundamental equations of the theory of elasticity in terms of harmonic functions [in Russian]. Izv. Akad. Nauk SSSR, Fiz.-Mat. Ser. **10**, 1425—1435.
1933 E 1. Marguerre, K.: Ebenes und achsensymmetrisches Problem der Elastizitätstheorie. Z. angew. Math. Mech. **13**, 437—438.
1934 E 1. Biezeno, C.: Über die Marguerresche Spannungsfunktion. Ing.-Archiv **5**, 120—124.
E 2. Galerkin, B.: Contribution to the theory of an elastic cylindrical shell. C. r. Doklady URSS (2) **4**, 270—275.
E 3. Neuber, H.: Ein neuer Ansatz zur Lösung räumlicher Probleme der Elastizitätstheorie. Der Hohlkegel unter Einzellast als Beispiel. Z. angew. Math. Mech. **14**, 203—212.

E 4. SOBRERO, L.: Nuovo metodo per lo studio dei problemi di elasticità con applicazione al problema della piastra forata. Ricerche di Ingegneria **2**, 255—264.

1935 E 1. SOBRERO, L.: Delle funzioni analoghe al potentiale intervenienti nella fisica-matematica. Rend. Lincei. (6) **21**, 448—454.

E 2. STEUERMANN, E.: Sur une transformation des équations de la théorie d'élasticité. J. inst. math. acad. sci. Ukraine **1935**, Nos. 3—4, 35—40.

E 3. WESTERGAARD, H.: General solution of the problem of elastostatics of an n-dimensional homogeneous isotropic solid in an n-dimensional space. Bull. Amer. math. soc. **41**, 695—699.

1936 E 1. MINDLIN, R.: Note on the GALERKIN and PAPKOVICH stress functions. Bull. Amer. math. soc. **42**, 373—376.

E 2. ZANABONI, O.: Il problema della funzione delle tensioni in un sistema spaziale isotropo. Bull. un. mat. Ital. **15**, 71—76.

1937 E 1. NEUBER, H.: Kerbspannungslehre. Berlin: Springer. VII + 160 pp. See § 6.

E 2. TOLOTTI, C.: Sui problemi di elasticità piana a funzione di AIRY polidroma. Rend. Lincei (6) **25**, 226—230.

1938 E 1. SLOBODIANSKI, M.: Stress functions for spatial problems in the theory of elasticity [in Russian]. Uchenie Zap. Moskov. Gos. Univ. **24**, 181—190.

E 2. SLOBODIANSKI, M.: Expression of the solution of the differential equations of elasticity by means of one, two, and three functions and proof of the generality of these solutions [in Russian]. Uchenie Zap. Moskov. Gos. Univ. **24**, 191—202.

1939 E 1. BIEZENO, C. B., & R. GRAMMEL: Technische Dynamik. Berlin: Springer. See Ch. II, § 3.

1940 E 1. LEHKNITSKY, S.: Symmetric deformation and torsion of bodies of revolution with anisotropy of a special kind [in Russian]. Priklad. math. mekh. (3) **4**, No. 3, 43—60.

E 2. LOCATELLI, P.: Sulla congruenza delle deformazioni. Rend. ist. Lombardo **73**, (4) 3, 457—464 (1939—1940).

1941 E 1. PLATRIER, C.: Sur l'intégration des équations indéfinies de l'équilibre élastique. C. r. acad. sci. Paris **212**, 749—751.

1942 E 1. SOUTHWELL, R.: Some practically important stress-systems in solids of revolution. Proc. r. soc. London A **180**, 367—396.

1944 E 1. CARRIER, G. F.: The thermal-stress and body-force problems for the infinite orthotropic solid. Q. appl. math. **2**, 31—36.

E 2. FÖPPL, A. & L.: Drang und Zwang 2. Berlin: Oldenbourg. See p. 207 *et sqq.*

1946 E 1. BUTTY, E.: Tratado de elasticidad teorico-tecnica **1**, Buenos Aires: Centro Estudiantes de Ingenieria. See Chs. VII—VIII.

1947 E 1. GRIOLI, G.: Struttura della funzione di AIRY nei sistemi molteplicemente connessi. Giorn. mat **77**, 119—144; trans., The structure of AIRY's stress function in multiply connected regions, NACA TM 1290, 1951.

E 2. GUTMAN, C. G.: General solution of a problem in the theory of elasticity in generalized cylindrical coordinates [in Russian]. Doklad. Akad. Nauk SSSR **58**, 993—996.

E 3. SADOWSKY, M., & E. STERNBERG: Stress concentration around an ellipsoidal cavity in an infinite body under arbitrary plane stress perpendicular to the axis of revolution of cavity. J. appl. mech. **69**, A 191—A 201.

E 4. SHAPIRO, G.: Les fonctions des tensions dans un système arbitraire de coordonnées curvilignes. C. r. Doklady acad. sci. URSS **55**, 693—695.

1948 E 1. AYMERICH, G.: Trasformazione conforme delle funzioni biarmoniche ed applicazione all' elasticità piana. Rend. sem. fac. sci. univ. Cagliari **17**, 1—12 (1947).

E 2. ELLIOT, H.: Three-dimensional stress distributions in hexagonal aeolotropic crystals. Proc. Cambridge phil. soc. **44**, 522—533.

E 3. MOISIL, GR. C.: Sur une généralisation de la fonction d'AIRY. Bull. éc. poly. Jassy **3**, 156.

E 4. Timpe, A.: Torsionsfreie achsensymmetrische Deformation von Umdrehungskörpern und ihre Inversion. Z. angew. Math. Mech. **28**, 161—166.

1949 E 1. Aquaro, G.: Sul calcolo delle deformazioni di uno strato sferico elastico. Rend. Lincei (8) **7**, 289—297.

E 2. Fichera, G.: Sul calcolo delle deformazioni, dotate di simmetria assiale, di uno strato sferico elastico. Rend. Lincei (8) **6**, 582—590.

E 3. Freiberger, W.: The uniform torsion of an incomplete tore. Austral. J. sci. A **2**, 354—375.

E 4. Freiberger, W.: On the solution of the equilibrium equations of elasticity in general curvilinear coordinates. Austral. J. sci. A **2**, 483—492.

E 5. Iacovache, M.: O extindere a metodei lui Galerkin pentru sistemul ecuaţiilor elasticităţii. Acad. Române. Bul. Şti. A **1**, 593—596.

E 6. Martin, F.: Die Membran-Kugelschale unter Einzellasten. Ing.-Arch. **17**, 167—186.

E 7. Moisil, G.: Asupra sistemelor de ecuaţii cu derivate parţiale lineare şi cu coefficienţi constanti. Acad. Române. Bul. Şti. A **1**, 341—351.

E 8. Moisil, G.: Asupra formelelor lui Galerkin în teoria elasticităţii. Acad. Române. Bul. Şti. A **1**, 587—592.

E 9. Moisil, G.: Un analog al vectorului lui Galerkin in hidrodinamica lichidelior vascoase. Acad. Române. Bul. Şti. A **1**, 803—812.

E 10. Sternberg, E., & M. Sadowsky: Three-dimensional solution for the stress concentration around a circular hole in a plate of arbitrary thickness. J. appl. mech. **16**, 27—38.

E 11. Ziemba, S.: Fonction des tensions aux coordonnées sphèriques dans le cas d'une symétrie axiale des déformations et des tensions. Archiwum Mech. Stosow **1**, 311—338.

1950 E 1. Ionescu-Cazimir, V.: Asupra ecuaţiilor echilibrului termoelastic. I. Analogul vectorului lui Galerkin. Acad. Române. Bul. Şti. Ser. Mat. Fiz. Chim. **2**, 589—595.

E 2. Lehknitsky, S.: Theory of Elasticity of an Anisotropic Body [in Russian]. Moscow-Leningrad.

E 3. Moisil, A.: Asupra unui vector analog vectorului lui Galerkin pentru echilibrul corpurilor elastice cu isotropie transversă. Acad. Române. Bul. Şti. Ser. Mat. Fiz. Chim. **2**, 207—210.

E 4. Timpe, A.: Spannungsfunktion für die von Kugel- und Kegelflächen begrenzten Körper und Kuppelproblem. Z. angew. Math. Mech. **30**, 50—61.

1951 E 1. Iacovache, M.: Relaţiile între tensiuni într'un lichid vâscos incompresibil în mişcare lentă, permanenta. Comun. acad. Române **1**, 245—249.

E 2. Ionescu-Cazimir, V.: Asupra ecuaţiilor echilibrului termoelastic. II. Relaţiile intre tensiuni şi temperatură. Comun. acad. Române **1**, 171—177.

E 3. Ionescu-Cazimir, V.: Asupra ecuaţiilor echilibrului termoelastic. III. Relaţiile intre tensiuni. Comun. acad. Române **1**, 385—390.

E 4. Koçо, P.: Asupra echilibrului unei clase de corpuri visco-elastice. Acad. Române. Bul. şti. sect. mat. fiz. **3**, 221—243.

E 5. Sternberg, E., R. Eubanks & M. Sadowsky: On the stress-function approaches of Boussinesq and Timpe to the axisymmetric problem of elasticity theory. J. appl. phys. **22**, 1121—1124.

E 6. Timpe, A.: Spannungsfunktion achsensymmetrischer Deformationen in Zylinderkoordinaten. Z. angew. Math. Mech. **31**, 220—224.

1952 E 1. Iacovache, M.: Aplicarea funcţiilor monogene in sensul lui Feodorov la teoria elasticitaţii corpurilor cu izotropie transversa. Rev. univ. „C. I. Parhon" politeh. ser. şti. Mat. **1**, No. 1, 58—60.

E 2. Ionescu-Cazimir, V.: Asupra ecuaţiilor echilibrului termoelastic plan. Rev. univ. „C. I. Parhon" politeh. ser. şti. Mat. **1**, No. 1, 55—57.

E 3. Ionescu-Cazimir, V.: Asupra ecuaţiilor echilibrului termoelastic. IV. Cazul plan. Acad. Române. Bul. şti. sect. mat. fiz. **4**, 547—554.

E 4. MALITA, M.: Asupra ecuațiilor de mișcare ale corpurilor elastice cu isotropie transversală. Comun. acad. Române **2**, 681—689.
E 5. MOISIL, G. C.: Teoria preliminară a sistemelor de ecuații cu derivate parțiale cu coeficienți constanți. Bul. științ. Acad. Pop. Române **4**, 319ff.
E 6. TEODORESCU, P.: Asupra teorili exacte a echilibrului suprafetelor cilindrice. Acad. Române. Bul. și. ser. mat. fiz. **4**, 111—193.
E 7. WESTERGAARD, H.: Theory of Elasticity and Plasticity. xii + 176 pp. Cambridge. See Ch. VI.

1953 E 1. ARZHANIKH, I.: Parametric representation of solutions of a system of linear functional equations in commutative operators [in Russian]. Uspekhi Mat. Nauk (2) **8**, No. 3 (55), 157—160.
E 2. ARZHANIKH, I.: Studies on the mechanics of continuous media [in Russian]. Akad. Nauk Uzbekh. SSR Trudi Inst. Mat. Mekh. **9**, 60—101.
E 3. BISHOP, R. E. D.: On dynamical problems of plane stress and plane strain. Q. J. Mech. Appl. Math. **6**, 250—254.
E 4. BRDIČKA, M.: Equations of compatibility and stress functions in tensor form [in Russian]. Czechosl. J. Phys. **3**, 36—52.
E 5. CHURIKOV, F.: On a form of the general solution of the equilibrium equations for the displacements in the theory of elasticity [in Russian]. Prikl. Mat. Mekh. **17**, 751—754.
E 6. HU, H.: On the three-dimensional problems of the theory of elasticity of a transversely isotropic body. Acta sci. Sinica **2**, 145—151.
E 7. SADOWSKY, M., & E. STERNBERG: Pure bending of an incomplete torus. J. appl. mech. **20**, 215—226.
E 8. TRENIN, S.: On the solutions of the equilibrium equations of an axisymmetrical problem in the theory of elasticity [in Russian]. Vestnik Moskov. Univ. Ser. Fiz.-Mat. **8**, 7—13.

1954 E 1. BRDIČKA, M.: Covariant form of the general solution of the GALERKIN equations of elastic equilibrium [in Russian]. Czechosl. J. Phys. **4**, 246.
E 2. EUBANKS, R., & E. STERNBERG: On the axisymmetric problem of elasticity theory for a medium with transverse isotropy. J. Rational Mech. Anal. **3**, 89—101.
E 3. HU, H.: On the general theory of elasticity for a spherically isotropic medium. Sci. Sinica **3**, 247—260.
E 4. IONESCU, D.: Asupra vectorului lui GALERKIN în teoria elasticității si în hidrodinamica fluidelor vîscoase. Acad. Române. Bul. și. sect. mat. fiz. **6**, 555—571.
E 5. KRÖNER, E.: Die Spannungsfunktionen der dreidimensionalen isotropen Elastizitätstheorie. Z. Physik **139**, 175—188. Correction, Z. Physik **143**, 175 (1955).
E 6. LING, C.-B., & K.-L. YANG: On symmetrical strains in solids of revolution in curvilinear coordinates. Ann. Acad. Sinica Taipei **1**, 507—516.
E 7. MINDLIN, R. D.: Force at a point in the interior of a semi-infinite solid. Proc. 1st Midwest Conf. Solid Mech. 1953, Univ. Illinois, 56—59.
E 8. SLOBODIANSKI, M. G.: The general form of solutions of the equations of elasticity for simply connected and multiply connected domains, expressed by harmonic functions [in Russian]. Prikl. Mat. Mekh. **18**, 55—74.

1955 E 1. MARGUERRE, K.: Ansätze zur Lösung der Grundgleichungen der Elastizitätstheorie. Z. angew. Math. Mech. **35**, 241—263.

1956 E 1. BIOT, M. A.: Thermo-elasticity and irreversible thermodynamics. J. Appl. Phys. **27**, 240—253.
E 2. EUBANKS, R. A., & E. STERNBERG: On the completeness of the BOUSSINESQ-PAPKOVICH stress functions. J. Rational Mech. Anal. **5**, 735—746.
E 3. RADOK, J. R. M.: On the solution of problems of dynamic plane elasticity. Q. Appl. Math. **14**, 289—298.
E 4. TEODORESCU, P. P.: On the plane problem of the elastodynamics. Rev. Méc. Appl. Acad. Rep. Pop. Roumaine **1**, 179ff.

E 5. Teodorescu, P.: On a general method of solving the plane problem of elastodynamics [in Romanian]. Com. Acad. R. P. Romîne **6**, 795—801.
1957 E 1. Noll, W.: Verschiebungsfunktionen für elastische Schwingungsprobleme. Z. angew. Math. Mech. **37**, 81—87.
E 2. Sternberg, E., & R. A. Eubanks: On stress functions for elastokinetics and the integration of the repeated wave equation. Q. Appl. Math. **15**, 149—153.
1958 E 1. Blokh, V. I.: On the representation of the general solution of the basic equations of the static theory of elasticity for an isotropic body with the aid of harmonic functions [in Russian]. Prikl. Mat. Mekh. **22**, 473—479. English. transl., PMM **22**, 659—668.
E 2. Deev, V. M.: The solution of the spatial problem in the theory of elasticity [in Ukrainian]. Dopov. Acad. Nauk Ukrain. RSR 1958, 29—32.
E 3. Deresiewicz, H.: Solution of the equations of thermoelasticity. Proc. Third U.S. Nat. Cong Appl. Mech.
E 4. Duffin, R. J., & W. Noll: On exterior boundary value problems in linear elasticity. Arch. Rational Mech. Anal. **2**, 191—196 (1958—1959).
E 5. Predeleanu, M.: Über die Verschiebungsfunktion für das achsensymmetrische Problem der Elastodynamik. Z. angew. Math. Mech. **38**, 402—405.
1960 E 1. Sternberg, E.: On the integration of the equations of motion in the classical theory of elasticity. Arch. Rational Mech. Anal. (in press).

801 North College Avenue
Bloomington, Indiana

(Received July 1, 1959)

Infinitesimal Plane Strain in a Network of Elastic Cords

S. M. Genensky & R. S. Rivlin

1. Introduction

In a previous paper*, the problem has been considered of plane strain in a sheet of fabric consisting of a network formed by two families of parallel, inextensible, perfectly flexible cords, which cannot move relative to each other at their points of intersection. No restrictions were placed on the magnitude of the displacement gradients in the deformed sheet, except insofar as they are implied by the inextensibility of the cords.

In the present paper, we consider a similar problem in the case when the cords are extensible and elastic and the displacement gradients are sufficiently small so that terms involving the displacement gradients linearly may be neglected in comparison with unity, terms of second degree in the displacement gradients may be neglected in comparison with those of first degree and so on.

In Part I, the theory is developed without placing any limitations on the magnitude of the deformation undergone by the sheet and equations of equilibrium and expressions for the edge tractions are obtained for plane strain of the sheet. In Part II the relations obtained in Part I are simplified for the case when the displacement gradients are small compared with unity. It is shown (in § 8) how, in this case, the displacements throughout the sheet may be obtained when the edge displacements are specified. In § 9 it is shown how the displacements throughout the sheet may be obtained when the edge tractions are specified. Finally, in § 10 it is shown how the displacements throughout the sheet may be obtained in certain problems in which the boundary conditions are of the mixed type, the edge displacements being specified over part of the boundary and the edge tractions over the remainder of the boundary.

I. General Theory

2. Kinematic Considerations

The plane sheet of fabric described in § 1 is considered to undergo a plane deformation. We may describe the deformation with reference to a fixed rectangular Cartesian coordinate system x, the axes x_1 and x_2 of which are parallel to the bisectors of the angles between the cords in the undeformed state of the sheet, and are chosen so that the two families of cords make angles $\pm\alpha$ with the axis x_1 in the undeformed state of the sheet.

* Rivlin, R. S.: J. Rational Mech. Anal. **4**, 951 (1955).

If X_i and x_i $(i=1, 2)$ * are the coordinates in the system x of a generic particle of the sheet in the undeformed and deformed states respectively, then the deformation of the sheet is described by relations of the form

$$x_i = x_i(X_j), \tag{2.1}$$

where the indicated functional dependence is single-valued, possesses a unique inverse, and is continuous and differentiable as many times as may be required, except possibly at isolated points or on isolated lines. Let P and Q be two neighboring particles of the sheet and let dS and ds be the distances between them in the undeformed and deformed states respectively. Then,

$$(dS^2) = dX_i\, dX_i \tag{2.2}$$

and

$$(ds)^2 = dx_i\, dx_i. \tag{2.3}$$

From (2.1), it follows that

$$dx_i = \frac{\partial x_i}{\partial X_j} dX_j. \tag{2.4}$$

Using (2.4), (2.3) becomes

$$(ds)^2 = \frac{\partial x_i}{\partial X_j} \frac{\partial x_i}{\partial X_k} dX_j\, dX_k. \tag{2.5}$$

Let L_i be the direction-cosines, in the rectangular Cartesian coordinate system x, of the line element PQ in the undeformed state of the sheet. Then,

$$L_i = dX_i/dS. \tag{2.6}$$

With (2.5), we obtain

$$\left(\frac{ds}{dS}\right)^2 = \frac{\partial x_i}{\partial X_j} \frac{\partial x_i}{\partial X_k} L_j L_k. \tag{2.7}$$

Let e_1 and e_2 be the fractional extensions, in the deformed state of the sheet, of cords which make angles α and $-\alpha$, respectively with the x_1 axis in the undeformed state. Then, taking $L_j = \cos\alpha$, $\sin\alpha$ in (2.7), we obtain

$$(1+e_1)^2 = \left(\frac{\partial x_i}{\partial X_1}\cos\alpha + \frac{\partial x_i}{\partial X_2}\sin\alpha\right)\left(\frac{\partial x_i}{\partial X_1}\cos\alpha + \frac{\partial x_i}{\partial X_2}\sin\alpha\right). \tag{2.8}$$

Similarly, taking $L_j = \cos\alpha$, $-\sin\alpha$ in (2.7), we obtain

$$(1+e_2)^2 = \left(\frac{\partial x_i}{\partial X_1}\cos\alpha - \frac{\partial x_i}{\partial X_2}\sin\alpha\right)\left(\frac{\partial x_i}{\partial X_1}\cos\alpha - \frac{\partial x_i}{\partial X_2}\sin\alpha\right). \tag{2.9}$$

Let Z_i denote the coordinates, in an oblique coordinate system, the axes of which are parallel to the directions of the two families of cords in the undeformed sheet, of a point which is at X_i in the system x. We then have

$$X_1 = (Z_1 + Z_2)\cos\alpha \quad \text{and} \quad X_2 = (Z_1 - Z_2)\sin\alpha. \tag{2.10}$$

From (2.10), we obtain

$$\begin{aligned} \frac{\partial}{\partial Z_1} &= \cos\alpha \frac{\partial}{\partial X_1} + \sin\alpha \frac{\partial}{\partial X_2} \\ \text{and} \qquad \frac{\partial}{\partial Z_2} &= \cos\alpha \frac{\partial}{\partial X_1} - \sin\alpha \frac{\partial}{\partial X_2}. \end{aligned} \tag{2.11}$$

* Latin subscripts will be considered to take the values 1, 2, throughout this paper.

Introducing (2.11) into (2.8) and (2.9), we obtain

$$(1+e_1)^2 = \frac{\partial x_i}{\partial Z_1}\frac{\partial x_i}{\partial Z_1} \quad \text{and} \quad (1+e_2)^2 = \frac{\partial x_i}{\partial Z_2}\frac{\partial x_i}{\partial Z_2}. \tag{2.12}$$

Let l_i be the direction-cosines of the line element PQ in the deformed state of the sheet, in the coordinate system x. Then

$$l_i = dx_i/ds. \tag{2.13}$$

With equations (2.4) and (2.6), we obtain

$$l_i = \frac{dS}{ds}\frac{\partial x_i}{\partial X_j}L_j. \tag{2.14}$$

3. *Stresses*

We consider next a sheet of fabric subjected to plane strain by the application of edge and surface tractions. The components of stress, t_{ij}, resulting from this deformation, are defined in the following manner; t_{i1} and t_{i2} are the components of the force per unit length in the positive directions of the x_1 and x_2 axes respectively, measured in the deformed state, exerted across an element of length at (x_1, x_2) normal to the x_i axis, by the material on the positive side of the element upon the material on the negative side of the element. From the continuity of the fabric and the equilibrium of moments, it follows that

$$t_{ij} = t_{ji}. \tag{3.1}$$

Let T_i be the components in the coordinate directions of the stress vector, measured in the deformed state, acting on an element of length at (x_1, x_2) with unit normal which has direction-cosines n_i in the system x. Then,

$$T_i = t_{ij}n_j. \tag{3.2}$$

4. *The Relation Between Stress and Deformation*

We assume that the sheet of fabric is deformed in accordance with equation (2.1) by forces acting in the plane of the sheet. Let τ_1 and τ_2 be the tensions, at a generic point P of the sheet, in the cords inclined initially at angles α and $-\alpha$ to the x_1 axis. Let β_1 and $-\beta_2$ be the inclinations of these cords to the x_1 axis in the deformed state.

Since linear elements of the cords of the two families, which have lengths dZ_1 and dZ_2 respectively in the undeformed state, have lengths $(1+e_1)\,dZ_1$ and $(1+e_2)\,dZ_2$ respectively in the deformed state, we have,

and

$$\begin{aligned} \cos\beta_1 &= \frac{1}{1+e_1}\frac{\partial x_1}{\partial Z_1}, \qquad \sin\beta_1 = \frac{1}{1+e_1}\frac{\partial x_2}{\partial Z_1}, \\ \cos\beta_2 &= \frac{1}{1+e_2}\frac{\partial x_1}{\partial Z_2}, \qquad \sin\beta_2 = -\frac{1}{1+e_2}\frac{\partial x_2}{\partial Z_2}. \end{aligned} \tag{4.1}$$

These results can be obtained by taking $L_1, L_2 = \cos\alpha, \sin\alpha$ and $l_1, l_2 = \cos\beta_1, \sin\beta_1$ and by taking $L_1, L_2 = \cos\alpha, -\sin\alpha$ and $l_1, l_2 = \cos\beta_2, -\sin\beta_2$ in equations (2.14) and employing the relations (2.11).

We denote by $\bar{d}$ the intercept formed by adjacent cords of one family on a cord of the other family in the undeformed state. Then the intercept formed, in the deformed state, by adjacent cords at P of the family initially parallel to the Z_2 axis, on the cord through P, initially parallel to the Z_1 axis, is $\bar{d}(1+e_1)$. The number of cords of the Z_2 family intersecting the cord of the Z_1 family per unit length in the deformed state is therefore $1/\bar{d}(1+e_1)$. The stress vector on the Z_1 cord per unit length is thus $\tau_2/\bar{d}(1+e_1)$ in a direction parallel to the cords of the Z_2 family, *i.e.* in a direction having direction-cosines $\cos\beta_2$, $-\sin\beta_2$ in the coordinate system x. The components of the stress vector in the direction of the x_1 and x_2 axes are therefore

$$\frac{\tau_2\cos\beta_2}{\bar{d}(1+e_1)} \quad \text{and} \quad -\frac{\tau_2\sin\beta_2}{\bar{d}(1+e_1)} \tag{4.2}$$

respectively. The normal to the element of the Z_1 cord considered has direction-cosines $(\sin\beta_1, -\cos\beta_1)$. From equations (3.2) we see that the stress vector acting on such an element has components

$$t_{11}\sin\beta_1 - t_{12}\cos\beta_1 \quad \text{and} \quad t_{12}\sin\beta_1 - t_{22}\cos\beta_1 \tag{4.3}$$

in the x_1 and x_2 directions respectively. We thus have from (4.2) and (4.3)

$$\begin{aligned} t_{11}\sin\beta_1 - t_{12}\cos\beta_1 &= \frac{\tau_2\cos\beta_2}{\bar{d}(1+e_1)} \\ \text{and} \qquad t_{12}\sin\beta_1 - t_{22}\cos\beta_1 &= -\frac{\tau_2\sin\beta_2}{\bar{d}(1+e_1)}. \end{aligned} \tag{4.4}$$

In a similar fashion, by considering the stress vector acting on an element of the Z_2 cord through P, we obtain

$$\begin{aligned} t_{11}\sin\beta_2 + t_{12}\cos\beta_2 &= \frac{\tau_1\cos\beta_1}{\bar{d}(1+e_2)} \\ \text{and} \qquad t_{12}\sin\beta_2 + t_{22}\cos\beta_2 &= \frac{\tau_1\sin\beta_1}{\bar{d}(1+e_2)}. \end{aligned} \tag{4.5}$$

Solving equations (4.4) and (4.5) for t_{11}, t_{22} and t_{12}, we obtain

$$\begin{aligned} t_{11} &= \frac{1}{\bar{d}\sin(\beta_1+\beta_2)}\left(\frac{\tau_1\cos^2\beta_1}{1+e_2} + \frac{\tau_2\cos^2\beta_2}{1+e_1}\right), \\ t_{22} &= \frac{1}{\bar{d}\sin(\beta_1+\beta_2)}\left(\frac{\tau_1\sin^2\beta_1}{1+e_2} + \frac{\tau_2\sin^2\beta_2}{1+e_1}\right) \\ \text{and} \qquad t_{12} &= \frac{1}{\bar{d}\sin(\beta_1+\beta_2)}\left(\frac{\tau_1\cos\beta_1\sin\beta_1}{1+e_2} - \frac{\tau_2\cos\beta_2\sin\beta_2}{1+e_1}\right). \end{aligned} \tag{4.6}$$

Let d be the distance between adjacent cords of each family in the undeformed state of the sheet. Then,

$$\bar{d} = d/\sin 2\alpha. \tag{4.7}$$

Substituting from (4.1) and (4.7) in (4.6), we obtain

$$\begin{aligned} t_{11} &= \frac{\sin 2\alpha}{d\Delta}\left[\frac{\tau_1}{1+e_1}\left(\frac{\partial x_1}{\partial Z_1}\right)^2 + \frac{\tau_2}{1+e_2}\left(\frac{\partial x_1}{\partial Z_2}\right)^2\right], \\ t_{22} &= \frac{\sin 2\alpha}{d\Delta}\left[\frac{\tau_1}{1+e_1}\left(\frac{\partial x_2}{\partial Z_1}\right)^2 + \frac{\tau_2}{1+e_2}\left(\frac{\partial x_2}{\partial Z_2}\right)^2\right] \\ \text{and} \qquad t_{12} &= \frac{\sin 2\alpha}{d\Delta}\left[\frac{\tau_1}{1+e_1}\frac{\partial x_1}{\partial Z_1}\frac{\partial x_2}{\partial Z_1} + \frac{\tau_2}{1+e_2}\frac{\partial x_1}{\partial Z_2}\frac{\partial x_2}{\partial Z_2}\right], \end{aligned} \tag{4.8}$$

where Δ is defined by

$$\Delta = \frac{\partial x_1}{\partial Z_2}\frac{\partial x_2}{\partial Z_1} - \frac{\partial x_1}{\partial Z_1}\frac{\partial x_2}{\partial Z_2}. \tag{4.9}$$

5. *The Equations of Motion*

The equations of motion for the sheet are

$$\frac{\partial t_{ij}}{\partial x_j} + F_i = \varrho \frac{\partial^2 x_i}{\partial t^2}, \tag{5.1}$$

where F_i denotes the applied force per unit area, measured in the deformed state, and ϱ denotes the mass of fabric per unit area, measured in the deformed state. We shall limit our analysis to static problems, so that $\partial^2 x_i/\partial t^2 = 0$ and equation (5.1) can then be written as

$$\frac{\partial t_{ij}}{\partial Z_k}\frac{\partial Z_k}{\partial x_j} + F_i = 0. \tag{5.2}$$

We have

$$\frac{\partial x_i}{\partial Z_k}\frac{\partial Z_k}{\partial x_j} = \delta_{ij}. \tag{5.3}$$

Regarding (5.3) as an equation for the determination of $\partial Z_k/\partial x_j$, we obtain the solution as

$$\frac{\partial Z_k}{\partial x_j} = \frac{1}{\Delta}\left(\frac{\partial x_k}{\partial Z_j} - \delta_{jk}\frac{\partial x_p}{\partial Z_p}\right), \tag{5.4}$$

where Δ is defined by equation (4.9).

Introducing (5.4) into (5.2), we obtain

$$\frac{\partial t_{ij}}{\partial Z_k}\left(\frac{\partial x_k}{\partial Z_j} - \delta_{jk}\frac{\partial x_p}{\partial Z_p}\right) + \Delta F_i = 0. \tag{5.5}$$

The expressions (4.8) may be introduced into (5.5) and the resulting equations simplified to yield

$$\begin{aligned} &\frac{\partial}{\partial Z_1}\left(\frac{\tau_1}{1+e_1}\frac{\partial x_1}{\partial Z_1}\right) + \frac{\partial}{\partial Z_2}\left(\frac{\tau_2}{1+e_2}\frac{\partial x_1}{\partial Z_2}\right) - \frac{d\,\Delta F_1}{\sin 2\alpha} = 0 \\ \text{and}\quad &\frac{\partial}{\partial Z_1}\left(\frac{\tau_1}{1+e_1}\frac{\partial x_2}{\partial Z_1}\right) + \frac{\partial}{\partial Z_2}\left(\frac{\tau_2}{1+e_2}\frac{\partial x_2}{\partial Z_2}\right) - \frac{d\,\Delta F_2}{\sin 2\alpha} = 0. \end{aligned} \tag{5.6}$$

If no surface forces act on the sheet, *i.e.* $F_1 = F_2 = 0$, equations (5.6) become

$$\begin{aligned} &\frac{\partial}{\partial Z_1}\left(\frac{\tau_1}{1+e_1}\frac{\partial x_1}{\partial Z_1}\right) + \frac{\partial}{\partial Z_2}\left(\frac{\tau_2}{1+e_2}\frac{\partial x_1}{\partial Z_2}\right) = 0 \\ \text{and}\quad &\frac{\partial}{\partial Z_1}\left(\frac{\tau_1}{1+e_1}\frac{\partial x_2}{\partial Z_1}\right) + \frac{\partial}{\partial Z_2}\left(\frac{\tau_2}{1+e_2}\frac{\partial x_2}{\partial Z_2}\right) = 0. \end{aligned} \tag{5.7}$$

6. *The Edge Tractions*

We now consider the plane sheet of fabric to be bounded by a simply-connected closed contour C, which is everywhere continuous and has a continuously-turning tangent except possibly at a finite number of points. We consider that the sheet is subjected to edge tractions in its plane, applied to the contour C.

Let Q be some reference particle and P a generic particle of the boundary C. We denote by Σ and σ the distance from Q to P, measured along C in the counter-clockwise sense, in the undeformed and deformed states of the sheet respectively. We further denote by N_i and n_i the direction-cosines of the outward-drawn normal to C at P in the undeformed and deformed states of the sheet respectively, and by M_i and m_i the direction-cosines of the tangent to C at P in these states. It follows immediately that

$$N_1 = M_2, \qquad N_2 = -M_1 \tag{6.1}$$

and

$$n_1 = m_2, \qquad n_2 = -m_1. \tag{6.2}$$

Let the components of the edge tractions, per unit length of contour, measured in the undeformed and deformed states of the sheet, be ξ_i and $\bar{\xi}_i$ respectively. Then,

$$\xi_i \, d\Sigma = \bar{\xi}_i \, d\sigma, \tag{6.3}$$

since $d\Sigma$ and $d\sigma$ are corresponding elements of length on the contour in the undeformed and deformed states respectively.

The components of edge traction $\bar{\xi}_i$ are given in terms of the stress components t_{ij} by

$$\bar{\xi}_i = t_{ij} n_j. \tag{6.4}$$

With (6.2), (6.4) yields

$$\bar{\xi}_1 = t_{11} m_2 - t_{12} m_1 \quad \text{and} \quad \bar{\xi}_2 = t_{12} m_2 - t_{22} m_1. \tag{6.5}$$

From (2.14), we obtain

$$m_i = \frac{d\Sigma}{d\sigma} \frac{\partial x_i}{\partial X_j} M_j. \tag{6.6}$$

Introducing (6.6) and (6.3) into (6.5), we obtain

$$\begin{aligned} \xi_1 &= \left(\frac{\partial x_2}{\partial X_j} t_{11} - \frac{\partial x_1}{\partial X_j} t_{12} \right) M_j \\ \text{and} \qquad \xi_2 &= \left(\frac{\partial x_2}{\partial X_j} t_{12} - \frac{\partial x_1}{\partial X_j} t_{22} \right) M_j. \end{aligned} \tag{6.7}$$

Introducing the expressions (4.8) for t_{ij} into (6.7), we obtain

$$\begin{aligned} \xi_1 = \frac{\sin 2\alpha}{d\Delta} \Bigg[& \frac{\tau_1}{1+e_1} \frac{\partial x_1}{\partial Z_1} \left(\frac{\partial x_1}{\partial Z_1} \frac{\partial x_2}{\partial X_j} - \frac{\partial x_2}{\partial Z_1} \frac{\partial x_1}{\partial X_j} \right) + \\ & + \frac{\tau_2}{1+e_2} \frac{\partial x_1}{\partial Z_2} \left(\frac{\partial x_1}{\partial Z_2} \frac{\partial x_2}{\partial X_j} - \frac{\partial x_2}{\partial Z_2} \frac{\partial x_1}{\partial X_j} \right) \Bigg] M_j \\ \text{and} \qquad \xi_2 = \frac{\sin 2\alpha}{d\Delta} \Bigg[& \frac{\tau_1}{1+e_1} \frac{\partial x_2}{\partial Z_1} \left(\frac{\partial x_1}{\partial Z_1} \frac{\partial x_2}{\partial X_j} - \frac{\partial x_2}{\partial Z_1} \frac{\partial x_1}{\partial X_j} \right) + \\ & + \frac{\tau_2}{1+e_2} \frac{\partial x_2}{\partial Z_2} \left(\frac{\partial x_1}{\partial Z_2} \frac{\partial x_2}{\partial X_j} - \frac{\partial x_2}{\partial Z_2} \frac{\partial x_1}{\partial X_j} \right) \Bigg] M_j. \end{aligned} \tag{6.8}$$

With (2.11) and (4.9), equations (6.8) yield

$$\begin{aligned} \xi_1 &= \frac{1}{d} \left[\frac{\tau_1}{1+e_1} \frac{\partial x_1}{\partial Z_1} (M_2 \cos\alpha - M_1 \sin\alpha) + \frac{\tau_2}{1+e_2} \frac{\partial x_1}{\partial Z_2} (M_2 \cos\alpha + M_1 \sin\alpha) \right] \\ \text{and} \qquad \xi_2 &= \frac{1}{d} \left[\frac{\tau_1}{1+e_1} \frac{\partial x_2}{\partial Z_1} (M_2 \cos\alpha - M_1 \sin\alpha) + \frac{\tau_2}{1+e_2} \frac{\partial x_2}{\partial Z_2} (M_2 \cos\alpha + M_1 \sin\alpha) \right]. \end{aligned} \tag{6.9}$$

II. Infinitesimal Deformations

7. Development of Basic Equations

Let u_i be the components in the coordinate system x of the displacement undergone in the deformation by a generic particle of the sheet. Then, we have

$$x_i = X_i + u_i. \tag{7.1}$$

We shall assume that the deformation is sufficiently small so that terms of the second degree in the displacement gradients $\partial u_i/\partial X_j$ may be neglected in comparison with those of the first degree. From (2.11), it is apparent that we may neglect terms of the second degree in $\partial u_i/\partial Z_j$ in comparison with those of the first degree. It is also apparent that we may neglect terms of the second degree in e_i in comparison with those of first degree.

With this approximation, we can write equations (2.12) as

$$\begin{aligned} &\cos\alpha \frac{\partial u_1}{\partial Z_1} + \sin\alpha \frac{\partial u_2}{\partial Z_1} = e_1 \\ \text{and} \quad &\cos\alpha \frac{\partial u_1}{\partial Z_2} - \sin\alpha \frac{\partial u_2}{\partial Z_2} = e_2. \end{aligned} \tag{7.2}$$

We shall assume that the fractional extensions e_1 and e_2 in the cords of the two families are related to the tensions τ_1 and τ_2 in them by the formulae

$$\tau_1 = k\, e_1 \quad \text{and} \quad \tau_2 = k\, e_2, \tag{7.3}$$

where k is a constant.

Employing the relations (7.1) and (7.3) in equations (5.7) and neglecting terms of the second degree in e_1, e_2 and $\partial u_i/\partial Z_j$, we obtain the equations of equilibrium, in the absence of surface forces as,

$$\begin{aligned} &\frac{\partial e_1}{\partial Z_1} + \frac{\partial e_2}{\partial Z_2} = 0 \\ \text{and} \quad &\frac{\partial e_1}{\partial Z_1} - \frac{\partial e_2}{\partial Z_2} = 0. \end{aligned} \tag{7.4}$$

From (7.4),

$$\frac{\partial e_1}{\partial Z_1} = \frac{\partial e_2}{\partial Z_2} = 0, \tag{7.5}$$

so that

$$e_1 = e_1(Z_2) \quad \text{and} \quad e_2 = e_2(Z_1); \tag{7.6}$$

i.e. the fractional extension—and hence the tension—is constant along each cord of the sheet, but may vary from cord to cord. Bearing this in mind, we now integrate equations (7.2) to obtain

$$\begin{aligned} &u_1 \cos\alpha + u_2 \sin\alpha = Z_1 e_1(Z_2) + p_1(Z_2) \\ \text{and} \quad &u_1 \cos\alpha - u_2 \sin\alpha = Z_2 e_2(Z_1) + p_2(Z_1), \end{aligned} \tag{7.7}$$

where $p_1(Z_2)$ and $p_2(Z_1)$ are arbitrary functions of their arguments. From (7.7), we obtain

$$\begin{aligned} &2u_1 \cos\alpha = Z_1 e_1(Z_2) + Z_2 e_2(Z_1) + p_1(Z_2) + p_2(Z_1) \\ \text{and} \quad &2u_2 \sin\alpha = Z_1 e_1(Z_2) - Z_2 e_2(Z_1) + p_1(Z_2) - p_2(Z_1). \end{aligned} \tag{7.8}$$

Again, employing the relations (7.1) and (7.3) in equations (4.8) and (4.9) and making the approximations employed above, we obtain the following expressions for the stress components t_{ij}:

$$t_{11} = \frac{k\cos^2\alpha}{d}(e_1+e_2), \qquad t_{22} = \frac{k\sin^2\alpha}{d}(e_1+e_2)$$

and

$$t_{12} = t_{21} = \frac{k\cos\alpha\sin\alpha}{d}(e_1-e_2). \tag{7.9}$$

In a similar manner, we obtain, from equations (6.9), the following expressions for the edge tractions ξ_i:

$$\xi_1 = \frac{k\cos\alpha}{d}\left[(e_1+e_2)M_2\cos\alpha - (e_1-e_2)M_1\sin\alpha\right]$$

and

$$\xi_2 = \frac{k\sin\alpha}{d}\left[(e_1-e_2)M_2\cos\alpha - (e_1+e_2)M_1\sin\alpha\right]. \tag{7.10}$$

8. Solution of the Problem when Edge Displacements are Specified

We consider a sheet which in the undeformed state is bounded by a simply-connected closed countour C which is everywhere continuous and has a continuously-turning tangent, except possibly at a finite number of points, as shown in Fig. 1. Let P be an interior point of the sheet with coordinates (Z_1, Z_2) in the oblique system formed by two intersecting cords. Two cords pass through P cutting C at four distinct point Q, R, S and T, which have coordinates (a_1, Z_2), (Z_1, a_2), (b_1, Z_2) and (Z_1, b_2) respectively.

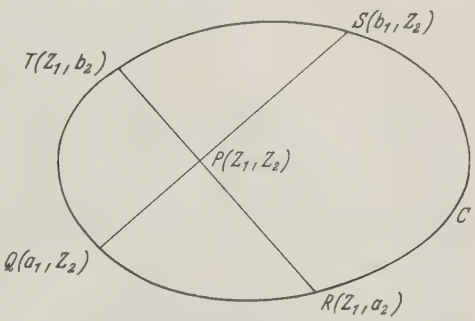

Fig. 1. Plane sheet bounded by simply-connected closed contour

From the first of equations (7.7), it follows that

$$\cos\alpha\, u_1(a_1, Z_2) + \sin\alpha\, u_2(a_1, Z_2) = a_1 e_1(Z_2) + p_1(Z_2) \text{ at } Q \tag{8.1}$$

and

$$\cos\alpha\, u_1(b_1, Z_2) + \sin\alpha\, u_2(b_1, Z_2) = b_1 e_1(Z_2) + p_1(Z_2) \text{ at } S. \tag{8.2}$$

Solving (8.1) and (8.2) for $e_1(Z_2)$ and $p_1(Z_2)$, we obtain

$$e_1(Z_2) = \frac{1}{a_1-b_1}\left\{\cos\alpha\left[u_1(a_1, Z_2) - u_1(b_1, Z_2)\right] + \sin\alpha\left[u_2(a_1, Z_2) - u_2(b_1, Z_2)\right]\right\}$$

and (8.3)

$$p_1(Z_2) = \frac{1}{b_1-a_1}\left\{\cos\alpha\left[b_1 u_1(a_1, Z_2) - a_1 u_1(b_1, Z_2)\right] + \sin\alpha\left[b_1 u_2(a_1, Z_2) - a_1 u_2(b_1, Z_2)\right]\right\}.$$

Again, from the second of equations (7.7), it follows that

$$\cos\alpha\, u_1(Z_1, a_2) - \sin\alpha\, u_2(Z_1, a_2) = a_2 e_2(Z_1) + p_2(Z_1) \text{ at } R \tag{8.4}$$

and

$$\cos\alpha\, u_1(Z_1, b_2) - \sin\alpha\, u_2(Z_1, b_2) = b_2 e_2(Z_1) + p_2(Z_1) \text{ at } T. \tag{8.5}$$

Solving (8.4) and (8.5) for $e_2(Z_1)$ and $p_2(Z_1)$, we obtain

$$e_2(Z_1) = \frac{1}{a_2 - b_2}\{\cos\alpha[u_1(Z_1, a_2) - u_1(Z_1, b_2)] - \sin\alpha[u_2(Z_1, a_2) - u_2(Z_1, b_2)]\}$$

and (8.6)

$$p_2(Z_1) = \frac{1}{b_2 - a_2}\{\cos\alpha[b_2 u_1(Z_1, a_2) - a_2 u_1(Z_1, b_2)] - \sin\alpha[b_2 u_2(Z_1, a_2) - a_2 u_2(Z_1, b_2)]\}.$$

Introducing (8.3) and (8.6) into (7.8), we arrive at

$$\begin{aligned} u_1(Z_1, Z_2) = {} & \frac{1}{2(a_1 - b_1)}\{(Z_1 - b_1)[u_1(a_1, Z_2) + \tan\alpha\, u_2(a_1, Z_2)] - \\ & - (Z_1 - a_1)[u_1(b_1, Z_2) + \tan\alpha\, u_2(b_1, Z_2)]\} + \\ & + \frac{1}{2(a_2 - b_2)}\{(Z_2 - b_2)[u_1(Z_1, a_2) - \tan\alpha\, u_2(Z_1, a_2)] - \\ & - (Z_2 - a_2)[u_1(Z_1, b_2) - \tan\alpha\, u_2(Z_1, b_2)]\} \end{aligned}$$

and (8.7)

$$\begin{aligned} u_2(Z_1, Z_2) = {} & \frac{1}{2(a_1 - b_1)}\{(Z_1 - b_1)[\cot\alpha\, u_1(a_1, Z_2) + u_2(a_1, Z_2)] - \\ & - (Z_1 - a_1)[\cot\alpha\, u_1(b_1, Z_2) + u_2(b_1, Z_2)]\} - \\ & - \frac{1}{2(a_2 - b_2)}\{(Z_2 - b_2)[\cot\alpha\, u_1(Z_1, a_2) - u_2(Z_1, a_2)] - \\ & - (Z_2 - a_2)[\cot\alpha\, u_1(Z_1, b_2) - u_2(Z_1, b_2)]\}. \end{aligned}$$

Thus if u_i are known on the entire boundary C, they are uniquely determined throughout the sheet by equations (8.7).

We shall now suppose that a portion AB of the contour C is formed by a cord $Z_2 = \text{const.}$, as shown in Fig. 2(i). It follows from the first of equations (7.7) that $u_1 \cos\alpha + u_2 \sin\alpha$ is a linear function of Z_1 on AB. Consequently, if u_1 is specified on AB and u_2 is specified at two points of it, u_2 can be determined over the whole of AB. Again, if u_2 is specified on AB and u_1 is specified at two points of it, u_1 can be determined over the whole of AB. It follows that if u_1 is specified over the whole of the contour C and u_2 is specified on the whole of

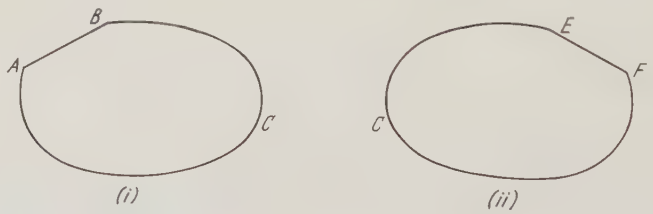

Fig. 2. Plane sheets bounded by simply-connected closed contour part of which is formed by a cord

the contour C except AB and at two points of AB (which may be the end-points A and B) then u_2 can be determined over the whole of the contour C and equations (8.7) used to determine u_1 and u_2 throughout the sheet. Alternatively, if u_2 is specified over the whole of the contour C and u_1 is specified over the whole of the contour C except AB and at two points of AB (which may be the end-points A and B) then u_1 can be determined over the whole of the contour C and equations (8.7) used to determine u_1 and u_2 throughout the sheet.

We now consider that a portion EF of the contour C is formed by a cord $Z_1 = \text{const}$, as shown in Fig. 2(ii). It follows from the second of equations (7.7) that $u_1 \cos\alpha - u_2 \sin\alpha$ is a linear function of Z_2 on EF. By an argument similar to that used in the previous case, we see that if one of the displacement components u_i is specified on the whole of the contour C and the other at all points of the contour C except EF and at two points of EF (which may be the end-points E and F), then both components of u_i can be determined over the whole of the contour C and equations (8.7) used to determine $u_i(Z_1, Z_2)$ throughout the sheet bounded by C.

It follows from this result that if the contour C is formed entirely by four cords, then if u_1 is specified at all points of C and u_2 at the vertices, u_2 may be determined over the whole of C. Alternatively, if u_2 is specified at all points of C and u_1 at the vertices, u_1 may be determined over the whole of C. In either case, equations (8.7) may then be used to determine u_1 and u_2 throughout the sheet.

9. *Solution of the Problem when Edge Tractions are Specified*

We consider the sheet of material shown in Fig. 1 and described in § 8. We assume that the edge traction ξ_i are specified over the whole boundary C of the sheet. Then, on C we have, solving equations (7.10) for $e_1(Z_2)$ and $e_2(Z_1)$,

$$e_1(Z_2) = \frac{\bar{d}\,[\xi_1(Z_1, Z_2)\sin\alpha + \xi_2(Z_1, Z_2)\cos\alpha]}{k\,[M_2(Z_1, Z_2)\cos\alpha - M_1(Z_1, Z_2)\sin\alpha]}$$

and

$$e_2(Z_1) = \frac{\bar{d}\,[\xi_1(Z_1, Z_2)\sin\alpha - \xi_2(Z_1, Z_2)\cos\alpha]}{k\,[M_2(Z_1, Z_2)\cos\alpha + M_1(Z_1, Z_2)\sin\alpha]}, \tag{9.1}$$

provided that

$$M_2(Z_1, Z_2)\cos\alpha \pm M_1(Z_1, Z_2)\sin\alpha \neq 0. \tag{9.2}$$

The condition (9.2) will, of course, be violated at points on C where the tangent to C is parallel to a cord direction.

If at a point on C the tangent to C is parallel to $Z_2 = \text{const}$, then $M_1, M_2 = \cos\alpha, \sin\alpha$. We then obtain, from (7.10) and (4.7),

$$e_2(Z_1) = \frac{\bar{d}}{k\cos\alpha}\,\xi_1(Z_1, Z_2) = -\frac{\bar{d}}{k\sin\alpha}\,\xi_2(Z_1, Z_2). \tag{9.3}$$

Again, if at a point on C the tangent to C is parallel to $Z_1 = \text{const}$, then $M_1, M_2 = \cos\alpha, -\sin\alpha$. We then obtain, from (7.10) and (4.7),

$$e_1(Z_2) = -\frac{\bar{d}}{k\cos\alpha}\,\xi_1(Z_1, Z_2) = -\frac{\bar{d}}{k\sin\alpha}\,\xi_2(Z_1, Z_2). \tag{9.4}$$

If the points Q, R, S and T are such that the condition (9.2) is not violated at any of them, then, from (9.1), we obtain

$$e_1(Z_2) = \frac{\bar{d}\,[\xi_1(a_1, Z_2)\sin\alpha + \xi_2(a_1, Z_2)\cos\alpha]}{k\,[M_2(a_1, Z_2)\cos\alpha - M_1(a_1, Z_2)\sin\alpha]} \quad \text{at } Q, \tag{9.5}$$

$$e_2(Z_1) = \frac{\bar{d}\,[\xi_1(Z_1, a_2)\sin\alpha - \xi_2(Z_1, a_2)\cos\alpha]}{k\,[M_2(Z_1, a_2)\cos\alpha + M_1(Z_1, a_2)\sin\alpha]} \quad \text{at } R, \tag{9.6}$$

$$e_1(Z_2) = \frac{\bar{d}\,[\xi_1(b_1, Z_2)\sin\alpha + \xi_2(b_1, Z_2)\cos\alpha]}{k\,[M_2(b_1, Z_2)\cos\alpha - M_1(b_1, Z_2)\sin\alpha]} \quad \text{at } S \tag{9.7}$$

and

$$e_2(Z_1) = \frac{\bar{d}\,[\xi_1(Z_1, b_2)\sin\alpha - \xi_2(Z_1, b_2)\cos\alpha]}{k\,[M_2(Z_1, b_2)\cos\alpha + M_1(Z_1, b_2)\sin\alpha]} \quad \text{at } T. \tag{9.8}$$

Comparing equations (9.5) and (9.7) and equations (9.6) and (9.8), we see that ξ_i cannot be arbitrarily specified on the boundary, but must satisfy the conditions

$$\text{and} \quad \begin{aligned} &\frac{\xi_1(a_1, Z_2)\sin\alpha + \xi_2(a_1, Z_2)\cos\alpha}{M_2(a_1, Z_2)\cos\alpha - M_1(a_1, Z_2)\sin\alpha} = \frac{\xi_1(b_1, Z_2)\sin\alpha + \xi_2(b_1, Z_2)\cos\alpha}{M_2(b_1, Z_2)\cos\alpha - M_1(b_1, Z_2)\sin\alpha} \\ &\frac{\xi_1(Z_1, a_2)\sin\alpha - \xi_2(Z_1, a_2)\cos\alpha}{M_2(Z_1, a_2)\cos\alpha + M_1(Z_1, a_2)\sin\alpha} = \frac{\xi_1(Z_1, b_2)\sin\alpha - \xi_2(Z_1, b_2)\cos\alpha}{M_2(Z_1, b_2)\cos\alpha + M_1(Z_1, b_2)\sin\alpha}. \end{aligned} \tag{9.9}$$

Introducing (9.5) and (9.6) into (7.8), we obtain

$$\text{and} \quad \begin{aligned} 2\cos\alpha\, u_1(Z_1, Z_2) = {} & Z_1 \frac{\bar{d}[\xi_1(a_1, Z_2)\sin\alpha + \xi_2(a_1, Z_2)\cos\alpha]}{k[M_2(a_1, Z_2)\cos\alpha - M_1(a_1, Z_2)\sin\alpha]} + \\ & + Z_2 \frac{\bar{d}[\xi_1(Z_1, a_2)\sin\alpha - \xi_2(Z_1, a_2)\cos\alpha]}{k[M_2(Z_1, a_2)\cos\alpha + M_1(Z_1, a_2)\sin\alpha]} + \\ & + p_1(Z_2) + p_2(Z_1) \\ 2\sin\alpha\, u_2(Z_1, Z_2) = {} & Z_1 \frac{\bar{d}[\xi_1(a_1, Z_2)\sin\alpha + \xi_2(a_1, Z_2)\cos\alpha]}{k[M_2(a_1, Z_2)\cos\alpha - M_1(a_1, Z_2)\sin\alpha]} - \\ & - Z_2 \frac{\bar{d}[\xi_1(Z_1, a_2)\sin\alpha - \xi_2(Z_1, a_2)\cos\alpha]}{k[M_2(Z_1, a_2)\cos\alpha + M_1(Z_1, a_2)\sin\alpha]} + \\ & + p_1(Z_2) - p_2(Z_1). \end{aligned} \tag{9.10}$$

We see that u_i are undetermined to the extent of a displacement field u_i' of the form

$$\text{and} \quad \begin{aligned} 2\cos\alpha\, u_1' &= p_1(Z_2) + p_2(Z_1) \\ 2\sin\alpha\, u_2' &= p_1(Z_2) - p_2(Z_1), \end{aligned} \tag{9.11}$$

which is an arbitrary (small) displacement field for which cord lengths remain unchanged. This indeterminacy of the displacement field u_i is not surprising in view of the fact that the fabric may be subjected to any deformation in which the lengths of cord elements are unchanged, without forces being applied.

If at one or more of the points Q, R, S, T, the tangent to C is parallel to $Z_2 = \text{const}$, we obtain $e_2(Z_1)$ from (9.3) and a knowledge of either ξ_1 or ξ_2 at that point. Similarly, if at one or more of the points Q, R, S, T, the tangent to C is parallel to $Z_1 = \text{const}$, we obtain $e_1(Z_2)$ from (9.4) and a knowledge of either ξ_1 or ξ_2 at that point. These relations may then be employed together with relations chosen from (9.5) to (9.8), appropriate to those of the points Q, R, S, T at which the tangent to C is not parallel to a cord, to provide expressions for $e_1(Z_2)$ and $e_2(Z_1)$ at P. These expressions may then be introduced into (7.8) to give expressions for $u_1(Z_1, Z_2)$ and $u_2(Z_1, Z_2)$ at P. By way of illustration, we consider below two cases.

(*i*) *Sheet bounded by cords.* We consider the boundary C of the sheet to be formed by the cords $Z_2 = a_2$, $Z_2 = b_2$, $Z_1 = a_1$, $Z_1 = b_1$, as shown in Fig. 3. Let P be a generic point (Z_1, Z_2) of the sheet and let Q, R, S, T be the intersections

with C of the cords through P. We then have, from (9.3) and (9.4),

$$\begin{aligned} e_2(Z_1) &= \frac{\bar{d}}{k\cos\alpha}\xi_1(Z_1, a_2) = -\frac{\bar{d}}{k\sin\alpha}\xi_2(Z_1, a_2) \quad \text{at } T, \\ e_2(Z_1) &= \frac{\bar{d}}{k\cos\alpha}\xi_1(Z_1, b_2) = -\frac{\bar{d}}{k\sin\alpha}\xi_2(Z_1, b_2) \quad \text{at } R, \\ e_1(Z_2) &= -\frac{\bar{d}}{k\cos\alpha}\xi_1(a_1, Z_2) = -\frac{\bar{d}}{k\sin\alpha}\xi_2(a_1, Z_2) \quad \text{at } Q, \\ e_1(Z_2) &= -\frac{\bar{d}}{k\cos\alpha}\xi(b_1, Z_2) = -\frac{\bar{d}}{k\sin\alpha}\xi_2(b_1, Z_2) \quad \text{at } S. \end{aligned} \tag{9.12}$$

Introducing (9.12) into (7.8), we obtain expressions for u_i with an arbitrariness corresponding to a small deformation of the sheet in which the lengths of cord elements are unaltered.

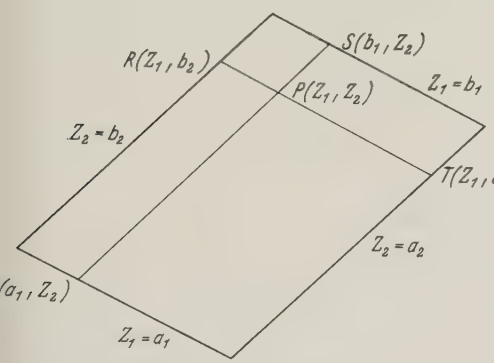

Fig. 3. Plane sheet bounded by four cords

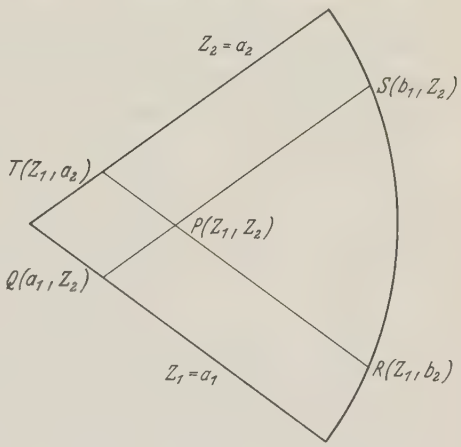

Fig. 4. Plane sheet bounded by two cords and an arc

(*ii*) *Sheet bounded by two intersecting cords and an arc.* We consider the boundary C of the sheet to be formed by the cords $Z_1 = a_1$ and $Z_2 = a_2$ and by an arc which is nowhere parallel to a cord, as shown in Fig. 4. Again P is a generic point of the sheet and Q, R, S, T are the intersections of the cords through P with the boundary. Suppose that S and R are the points (b_1, Z_2) and (Z_1, b_2) respectively.

From (9.3) and (9.4), we have

$$\begin{aligned} e_2(Z_1) &= \frac{\bar{d}}{k\cos\alpha}\xi_1(Z_1, a_2) = -\frac{\bar{d}}{k\sin\alpha}\xi_2(Z_1, a_2) \quad \text{at } T, \\ \text{and} \qquad e_1(Z_2) &= -\frac{\bar{d}}{k\cos\alpha}\xi_1(a_1, Z_2) = -\frac{\bar{d}}{k\sin\alpha}\xi_2(a_1, Z_2) \quad \text{at } Q. \end{aligned} \tag{9.13}$$

From (9.1), we have

$$\begin{aligned} e_1(Z_2) &= \frac{\bar{d}[\xi_1(b_1, Z_2)\sin\alpha + \xi_2(b_1, Z_2)\cos\alpha]}{k[M_2(b_1, Z_2)\cos\alpha - M_1(b_1, Z_2)\sin\alpha]} \quad \text{at } S \\ \text{and} \qquad e_2(Z_1) &= \frac{\bar{d}[\xi_1(Z_1, b_2)\sin\alpha - \xi_2(Z_1, b_2)\cos\alpha]}{k[M_2(Z_1, b_2)\cos\alpha + M_1(Z_1, b_2)\sin\alpha]} \quad \text{at } R. \end{aligned} \tag{9.14}$$

Employing either (9.13) or (9.14) in (7.8), we can obtain the displacement u_i throughout the sheet, with an arbitrariness corresponding to any deformation in which lengths of cord elements are unchanged.

We note that if the edge tractions are specified on the bounding cords, then it is sufficient that only one of the components ξ_i be specified on each of these cords. On the other hand, if the edge traction is specified on the bounding arc then both of the components ξ_i must be specified.

10. Mixed Boundary Value Problems

In this section we shall discuss a few examples which illustrate the manner in which the results obtained in the previous sections can be used to determine the displacements throughout a sheet when the boundary conditions are of the mixed type, *i.e.* displacements are given on parts of the boundary and edge tractions on other parts.

(*i*) *Sheet bounded by cords.* We consider the sheet of material bounded by four cords as shown in Fig. 3, and assume that the displacement component u_1 is specified on the edges $Z_1 - b_1$ and $Z_2 = b_2$, while the component of edge traction ξ_1 is specified on $Z_1 = a_1$ and $Z_2 = a_2$. From (9.3) and (9.4), we have

$$\text{and} \qquad \begin{aligned} e_2(Z_1) &= \frac{\bar{d}}{k\cos\alpha}\,\xi_1(Z_1, a_2) \quad \text{at } T \\ e_1(Z_2) &= -\frac{\bar{d}}{k\cos\alpha}\,\xi_1(a_1, Z_2) \quad \text{at } Q. \end{aligned} \tag{10.1}$$

From the first of equations (7.8), we obtain

$$\text{and} \qquad \begin{aligned} 2\cos\alpha\, u_1(b_1, Z_2) &= b_1 e_1(Z_2) + Z_2 e_2(b_1) + p_1(Z_2) + p_2(b_1) \quad \text{at } S, \\ 2\cos\alpha\, u_1(Z_1, b_2) &= Z_1 e_1(b_2) + b_2 e_2(Z_1) + p_1(b_2) + p_2(Z_1) \quad \text{at } R \\ 2\cos\alpha\, u_1(b_1, b_2) &= b_1 e_1(b_2) + b_2 e_2(b_1) + p_1(b_2) + p_2(b_1) \quad \text{at } Z_1 = b_1,\ Z_2 = b_2. \end{aligned} \tag{10.2}$$

Employing the first two of equations (10.2) to eliminate $p_1(Z_2)$ and $p_2(Z_1)$ from (7.8), we obtain

$$\text{and} \qquad \begin{aligned} 2\cos\alpha\, u_1(Z_1, Z_2) &= 2\cos\alpha\,[u_1(Z_1, b_2) + u_1(b_1, Z_2)] + \\ &\quad + (Z_1 - b_1)\, e_1(Z_2) + (Z_2 - b_2)\, e_2(Z_1) - \\ &\quad - Z_1 e_1(b_2) - Z_2 e_2(b_1) - p_1(b_2) - p_2(b_1) \\ 2\sin\alpha\, u_2(Z_1, Z_2) &= 2\cos\alpha\,[u_1(b_1, Z_2) - u_1(Z_1, b_2)] + \\ &\quad + (Z_1 - b_1)\, e_1(Z_2) - (Z_2 - b_2)\, e_2(Z_1) + \\ &\quad + Z_1 e_1(b_2) - Z_2 e_2(b_1) + p_1(b_2) - p_2(b_1). \end{aligned} \tag{10.3}$$

Eliminating $p_1(b_2) + p_2(b_1)$ from (10.3) and the third of equations (10.2), we obtain

$$\begin{aligned} 2\cos\alpha\, u_1(Z_1, Z_2) &= 2\cos\alpha\,[u_1(b_1, Z_2) + u_1(Z_1, b_2) - u_1(b_1, b_2)] + \\ &\quad + (Z_1 - b_1)\,[e_1(Z_2) - e_1(b_2)] + \\ &\quad + (Z_2 - b_2)\,[e_2(Z_1) - e_2(b_1)]. \end{aligned} \tag{10.4}$$

From (10.1), we have

$$\text{and} \qquad \begin{aligned} e_2(b_1) &= \frac{\bar{d}}{k\cos\alpha}\,\xi_1(b_1, a_2) \\ e_1(b_2) &= -\frac{\bar{d}}{k\cos\alpha}\,\xi_1(a_1, b_2). \end{aligned} \tag{10.5}$$

Introducing (10.1) and (10.5) into (10.4) and the second of equations (10.3), we obtain

$$\begin{aligned} u_1(Z_1,Z_2) &= u_1(b_1,Z_2)+u_1(Z_1,b_2)-u_1(b_1,b_2)- \\ &\quad -\frac{\bar{d}}{2k\cos^2\alpha}\{(Z_1-b_1)\,[\xi_1(a_1,Z_2)-\xi_1(a_1,b_2)]- \\ &\quad -(Z_2-b_2)\,[\xi_1(Z_1,a_2)-\xi_1(b_1,a_2)]\} \\ \text{and}\qquad u_2(Z_1,Z_2) &= \cot\alpha\,[u_1(b_1,Z_2)-u_1(Z_1,b_2)]- \\ &\quad -\frac{\bar{d}}{k\sin 2\alpha}\,[(Z_1-b_1)\,\xi_1(a_1,Z_2)+(Z_2-b_2)\,\xi_1(Z_1,a_2)+ \\ &\quad +Z_1\xi_1(a_1,b_2)+Z_2\xi_1(b_1,a_2)]+\frac{p_1(b_2)-p_2(b_1)}{2\sin\alpha}. \end{aligned} \tag{10.6}$$

Thus, if u_1 is specified on $Z_1=b_1$ and $Z_2=b_2$ and ξ_1 is specified on $Z_1=a_1$ and $Z_2=a_2$, then u_1 is determined throughout the sheet and u_2 is determined throughout the sheet apart from a constant. If, further, u_2 is specified at a single point of the sheet, then u_2 is completely determined throughout the sheet.

(*ii*) *Sheet bounded by two cords and an arc. Prescribed tractions on the arc and displacements on the cords.* We consider a sheet of material bounded by the cords $Z_1=a_1$ and $Z_2=a_2$ and by an arc which is nowhere parallel to a cord, as shown in Fig. 4. We assume that u_1 is specified on the cords $Z_1=a_1$ and $Z_2=a_2$ and that the edge traction ξ_i is specified on the arc.

Then, from the first of equations (7.8), we obtain

$$\begin{aligned} &2\cos\alpha\,u_1(Z_1,a_2)=Z_1e_1(a_2)+a_2e_2(Z_1)+p_1(a_2)+p_2(Z_1) \quad \text{at } T, \\ \text{and}\quad &2\cos\alpha\,u_1(a_1,Z_2)=a_1e_1(Z_2)+Z_2e_2(a_1)+p_1(Z_2)+p_2(a_1) \quad \text{at } Q \\ &2\cos\alpha\,u_1(a_1,a_2)=a_1e_1(a_2)+a_2e_2(a_1)+p_1(a_2)+p_2(a_1) \quad \text{at } Z_1=a_1,\ Z_2=a_2. \end{aligned} \tag{10.7}$$

From (10.7) and (7.8), we obtain, in a manner similar to that used in discussing case (*i*),

$$\begin{aligned} 2\cos\alpha\,u_1(Z_1,Z_2) &= 2\cos\alpha\,[u_1(a_1,Z_2)+u_1(Z_1,a_2)-u_1(a_1,a_2)]+ \\ &\quad +(Z_1-a_1)\,[e_1(Z_2)-e_1(a_2)]+ \\ &\quad +(Z_2-a_2)\,[e_2(Z_1)-e_2(a_1)] \\ \text{and}\qquad 2\sin\alpha\,u_2(Z_1,Z_2) &= 2\cos\alpha\,[u_1(a_1,Z_2)-u_1(Z_1,a_2)]+ \\ &\quad +(Z_1-a_1)\,e_1(Z_2)-(Z_2-a_2)\,e_2(Z_1)+ \\ &\quad +Z_1e_1(a_2)-Z_2e_2(a_1)+p_1(a_2)-p_2(a_1). \end{aligned} \tag{10.8}$$

Now, solving equations (7.10) for $e_1(Z_2)$ and $e_2(Z_1)$ and employing (4.7), we obtain

$$\begin{aligned} e_1(Z_2) &= \frac{\bar{d}\,[\xi_1(Z_1,Z_2)\sin\alpha+\xi_2(Z_1,Z_2)\cos\alpha]}{k\,[M_2(Z_1,Z_2)\cos\alpha-M_1(Z_1,Z_2)\sin\alpha]} \\ \text{and}\qquad e_2(Z_1) &= \frac{\bar{d}\,[\xi_1(Z_1,Z_2)\sin\alpha-\xi_2(Z_1,Z_2)\cos\alpha]}{k\,[M_2(Z_1,Z_2)\cos\alpha+M_1(Z_1,Z_2)\sin\alpha]}. \end{aligned} \tag{10.9}$$

In (10.9), the denominators are not zero on the bounding arc, since the arc is nowhere parallel to a cord. From (10.9) we have

$$\begin{aligned} e_1(Z_2) &= \frac{\bar{d}\,[\xi_1(b_1,Z_2)\sin\alpha+\xi_2(b_1,Z_2)\cos\alpha]}{k\,[M_2(b_1,Z_2)\cos\alpha-M_1(b_1,Z_2)\sin\alpha]} \quad \text{at } S \\ \text{and}\qquad e_2(Z_1) &= \frac{\bar{d}\,[\xi_1(Z_1,b_2)\sin\alpha-\xi_2(Z_1,b_2)\cos\alpha]}{k\,[M_2(Z_1,b_2)\cos\alpha+M_1(Z_1,b_2)\sin\alpha]} \quad \text{at } R. \end{aligned} \tag{10.10}$$

If the points of intersection of $Z_1=a_1$ and $Z_2=a_2$ with the arc are denoted (a_1, c_2) and (c_1, a_2) respectively, then from (10.9) we obtain

$$\text{and} \qquad \begin{aligned} e_1(a_2) &= \frac{\bar{d}\,[\xi_1(c_1, a_2)\sin\alpha + \xi_2(c_1, a_2)\cos\alpha]}{k\,[M_2(c_1, a_2)\cos\alpha - M_1(c_1, a_2)\sin\alpha]} \\ e_2(a_1) &= \frac{\bar{d}\,[\xi_1(a_1, c_2)\sin\alpha - \xi_2(a_1, c_2)\cos\alpha]}{k\,[M_2(a_1, c_2)\cos\alpha + M_2(a_1, c_2)\sin\alpha]}\,. \end{aligned} \tag{10.11}$$

Employing equations (10.10) and (10.11) to substitute for $e_1(Z_2)$, $e_2(Z_1)$, $e_1(a_2)$ and $e_2(a_1)$ in (10.8), we obtain expressions for $u_i(Z_1, Z_2)$ in terms of the known value of u_1 on the bounding cords $Z_1=a_1$ and $Z_2=a_2$ and the known values of the edge tractions ξ_i on the bounding arc. We note that u_2 is determined to within an arbitrary constant. However, if u_2 is prescribed at a single point of the sheet, it is then completely determined throughout the sheet.

(*iii*) *Sheet bounded by two cords and an arc. Prescribed tractions on the cords and displacements on the arc.* We again consider the sheet shown in Fig. 4, but we now assume that the displacement components u_i are known on the arc and the component of edge traction ξ_1 is known on the cords $Z_1=a_1$ and $Z_2=a_2$.

From (9.3) and (9.4) we obtain

$$\text{and} \qquad \begin{aligned} e_2(Z_1) &= -\frac{\bar{d}}{k\cos\alpha}\,\xi_1(Z_1, a_2) \quad \text{at } T \\ e_1(Z_2) &= -\frac{\bar{d}}{k\cos\alpha}\,\xi_1(a_1, Z_2) \quad \text{at } Q\,. \end{aligned} \tag{10.12}$$

From (7.7) it is seen that

$$\text{and} \qquad \begin{aligned} \cos\alpha\, u_1(Z_1, b_2) - \sin\alpha\, u_2(Z_1, b_2) &= b_2 e_2(Z_1) + p_2(Z_1) \quad \text{at } R \\ \cos\alpha\, u_1(b_1, Z_2) + \sin\alpha\, u_2(b_1, Z_2) &= b_1 e_1(Z_2) + p_1(Z_2) \quad \text{at } S\,. \end{aligned} \tag{10.13}$$

Employing (10.13) to eliminate $p_1(Z_2)$ and $p_2(Z_1)$ from (7.8), we have

$$\text{and} \qquad \begin{aligned} 2\cos\alpha\, u_1(Z_1, Z_2) &= \cos\alpha\,[u_1(b_1, Z_2) + u_1(Z_1, b_2)] + \\ &\quad + \sin\alpha\,[u_2(b_1, Z_2) - u_2(Z_1, b_2)] + \\ &\quad + (Z_1 - b_1)\, e_1(Z_2) + (Z_2 - b_2)\, e_2(Z_1) \\ 2\sin\alpha\, u_2(Z_1, Z_2) &= \cos\alpha\,[u_1(b_1, Z_2) - u_1(Z_1, b_2)] + \\ &\quad + \sin\alpha\,[u_2(b_1, Z_2) + u_2(Z_1, b_2)] + \\ &\quad + (Z_1 - b_1)\, e_1(Z_2) - (Z_2 - b_2)\, e_2(Z_1)\,. \end{aligned} \tag{10.14}$$

Substituting for $e_2(Z_1)$ and $e_1(Z_2)$ from (10.12) in (10.14) we obtain expressions for $u_i(Z_1, Z_2)$ in terms of the known component of edge traction ξ_1 on $Z_1=a_1$ and $Z_2=a_2$ and the known displacement components on the bounding arc.

Acknowledgement. The results presented in this paper were obtained in the course of research sponsored by the Office of Ordnance Research, U. S. Army, under Contract No. DA-19-020-ORD-4725 and its precursors.

Brown University
Providence, Rhode Island

(Received May 15, 1959)

Iterationsverfahren mit veränderlichen Operatoren

HANS EHRMANN

Vorgelegt von L. COLLATZ

§ 1. Einleitung*

Iterationsverfahren der Gestalt

$$u_{n+1} = T u_n, \qquad n = 0, 1, 2, \ldots, \tag{1.1}$$

zur Lösung einer Gleichung

$$u = T u \quad \text{oder} \quad S u = \Theta, \qquad (\Theta = \text{Nullelement}), \tag{1.2}$$

wobei die Größen $u, u_n, n = 0, 1, 2, \ldots$, Elemente eines abstrakten Raumes (z.B. eines Banachraumes) sind und T einen Operator bedeutet, sind in den letzten Jahren viel in bezug auf Konvergenzfragen, Konvergenzverbesserungen, Fehlerabschätzungen usw. untersucht worden**. Dabei zeigte es sich, daß viele bei speziellen Gleichungstypen bekannte Konvergenzkriterien und Fehlerabschätzungen, die vielfach in jedem speziellen Falle neu bewiesen wurden, Spezialfälle von z.T. sehr einfachen und leicht zu beweisenden abstrakten Sätzen sind und daß sich die Verschiedenheit von speziellen Konvergenzbedingungen und Fehlerabschätzungen aus der verschiedenen speziellen Wahl der Metrik ergibt [*3*], [*5*], [*10*]. In den meisten Fällen beruht die Konvergenz bei den betr. Sätzen auf der Eigenschaft des Operators T zu kontrahieren, und im einzelnen wird gewöhnlich gezeigt, daß die Summe der Abstände von u_{n+1} und u_n für $n = 0, 1, 2, \ldots$ konvergiert, was i. allg. mit Hilfe des Majorantenprinzips mittels einer Vergleichsreihe mit positiven Elementen geschieht, deren Konvergenz bekannt ist. Durch die abstrakte Formulierung der Sätze gelingt es zunächst,

---

* Dies ist ein erster Auszug einer Arbeit „Untersuchungen von Iterationsverfahren zur Lösung allgemeiner Gleichungen", die im November 1958 von der Fakultät für Natur- und Geisteswissenschaften der Technischen Hochschule Stuttgart als Habilitationsschrift anerkannt wurde. Die Untersuchungen entstanden während meiner Tätigkeit als Assistent am Institut für Mathematik und Mechanik der Bergakademie Clausthal bei Herrn Professor H. KÖNIG und am Mathematischen Institut A der Technischen Hochschule Stuttgart bei Herrn Professor G. SCHULZ. Daneben hat Herr Professor L. COLLATZ die Arbeit verfolgt und mich durch einige wertvolle Ratschläge und Hinweise unterstützt. Allen drei Professoren möchte ich für ihre freundliche Unterstützung herzlich danken.

** Siehe u. a. das Literaturverzeichnis am Ende der Arbeit. Hierauf beziehen sich die Literaturangaben in eckigen Klammern.

einen wesentlich größeren Kreis von Gleichungstypen (Gleichungssysteme, Differential- und Integralgleichungen usw.) zu erfassen. In der weiteren Entwicklung wurde sodann versucht, Konvergenzbedingungen und Fehlerabschätzungen der betr. abstrakten Sätze zu verbessern, indem die Definitionen insbes. der Metrik weiter verallgemeinert wurden, um die Abschätzungen dem jeweiligen speziellen Problem besser anpassen zu können. Dabei werden auch metrische Räume mit „Abständen" ihrer Punkte zugelassen, die keine nichtnegativen reellen Zahlen sind, [2], [7], [9]. Solche allgemeineren Räume, die den Banachraum als Spezialfall enthalten, werden wir auch hier zugrunde legen*.

Es treten nun Fälle auf, in denen z.B. die Iteration (1.1) noch sehr mühsam ist, so daß man zweckmäßig wenigstens in den ersten Schritten T durch einen Näherungsoperator $\tilde{T}$ ersetzt, oder T ist von vorneherein nur näherungsweise gegeben etwa in Form einer unendlichen Reihe, von der in (1.1) jeweils nur endlich viele Glieder berücksichtigt werden können, oder die Lösung u erscheint in Form einer unendlichen Reihe wie z.B. bei dem bekannten Verfahren zur Auflösung einer Gleichung der Form $x=f(x,\lambda)$, wobei man f an einer Stelle x_0 nach Potenzen von λ entwickelt, die Reihe abbricht, rechts für x einsetzt, wieder entwickelt, abbricht, einsetzt usw.:

$$x_{n+1}=f(x_n,\lambda)-\sum_{\nu\geq k_n} a_{\nu n}\lambda^\nu \quad \text{mit} \quad f(x_n,\lambda)=\sum_{\nu=0}^{\infty} a_{\nu n}\lambda^\nu.$$

Diese Fragestellungen führen auf das Problem, Iterationsverfahren der Form

$$u_{n+1}=T_n u_n \tag{1.3}$$

zu untersuchen, wobei sich also die Operatoren T_n von Schritt zu Schritt ändern können. Hiermit beschäftigt sich die vorliegende Arbeit. Es wird ein abstrakter Satz aufgestellt und bewiesen, der Bedingungen angibt, unter denen die Iteration (1.3) gegen eine Lösung der Gleichung (1.1) konvergiert. Außerdem wird eine Fehlerabschätzung angegeben, die unter den angegebenen Bedingungen nicht mehr verbessert werden kann. Der entsprechende Satz geht in dem Spezialfall $T_n=T$, $n=1,2,\ldots$, in bekannte Sätze über [4], [9]. Jedoch versagt hier die übliche Beweismethode. So braucht z.B. unter den getroffenen Voraussetzungen die Summe der Abstände von den aufeinanderfolgenden Iterationspunkten u_{n+1} und u_n $(n=0,1,2,\ldots)$ nicht zu konvergieren, noch kann man zeigen, daß diese Abstände schließlich monoton abnehmen. Obwohl der Satz dem Inhalt nach ganz elementar ist, ist er meines Wissens auch im einfachsten Fall einer unbekannten Zahl noch nicht in dieser Form mit Konvergenzbeweis und Fehlerabschätzung bekannt. Andererseits ist die hier durchgeführte abstrakte Formulierung so allgemein, daß der Satz auf sehr verschiedenartige Gleichungen leicht angewandt werden kann.

In § 2 werden alle funktionalanalytischen Hilfsmittel zusammengestellt. § 3 bringt den Satz und Beweis sowie einige für die Anwendungen auf den Nachweis der Existenz von Lösungen und ihre Abschätzungen wichtige Folgerungen. In § 4 werden schließlich einige Beispiele durchgeführt.

* Vgl. insbes. J. SCHRÖDER [9].

§ 2. Funktionalanalytische Voraussetzungen

Für je zwei Elemente $u, v \in R$ sei ein Abstand $\varrho(u, v)$ definiert, wobei die Abstände $\varrho, \sigma, \ldots$ Elemente eines linearen halbgeordneten Raumes N sind: es ist in N eine Addition $\varrho+\sigma$ und eine Multiplikation $\alpha\,\varrho$ mit reellen Zahlen α definiert, und diese Operationen genügen den üblichen Rechenregeln der Vektoralgebra. Ferner ist N halbgeordnet, d.h. es sind gewisse Elemente $\varrho, \sigma \in N$ als positiv definiert: $\varrho \geqq 0$ (0 = Nullelement von N), wobei $\varrho \geqq 0$ und $-\varrho \geqq 0$ (gleichzeitig) nur für $\varrho = 0$ gilt und $\varrho+\sigma \geqq 0$ aus $\varrho, \sigma \geqq 0$, sowie $\alpha\,\varrho \geqq 0$ aus $\alpha \geqq 0$ (α reelle Zahl) und $\varrho \geqq 0$ folgt. $\varrho \geqq \sigma$ oder $\sigma \leqq \varrho$ bedeutet $\varrho - \sigma \geqq 0$.

Ferner ist gewissen (konvergent genannten) Folgen $\varrho_n \in N$ ($n = 1, 2, 3, \ldots$) ein (eindeutig bestimmtes) Grenzelement $\lim \varrho_n \in N$ zugeordnet mit den Eigenschaften*:

(2.1)
- a) aus $\varrho_n = \varrho$ für $n = 1, 2, 3, \ldots$ mit einem festen Element ϱ folgt $\lim \varrho_n = \varrho$,
- b) aus $\lim \varrho_n = \varrho$ folgt $\lim \varrho_{k_n} = \varrho$ für jede monotone Folge natürlicher Zahlen k_n ($n = 1, 2, 3, \ldots$),
- c) aus $\lim \varrho_n = \varrho$ und $\lim \sigma_n = \sigma$ folgt $\lim(\varrho_n + \sigma_n) = \varrho + \sigma$,
- d) aus $\lim \alpha_n = \alpha$ (α_n, α reelle Zahlen) und $\lim \varrho_n = \varrho$ folgt $\lim \alpha_n \varrho_n = \alpha\varrho$,
- e) aus $0 \leqq \varrho_n \leqq \sigma_n$ für $n = 1, 2, 3, \ldots$ und $\lim \sigma_n = 0$ folgt $\lim \varrho_n = 0$,
- f) aus $\varrho_n \geqq 0$ für $n = 1, 2, 3, \ldots$ und $\lim \varrho_n = \varrho$ folgt $\varrho \geqq 0$,
- g) aus $0 \leqq \varrho_n \leqq \tau$ für $n \geqq n_0$ und $\lim \varrho_n = 0$ folgt die Existenz einer Nummer n_1 mit $\varrho_n \leqq \frac{\tau}{2}$ für $n \geqq n_1$.

Ferner gelten für den *bezüglich* N *metrischen Raum* R die Abstandspostulate:

1. $\varrho(u, v) = 0$ genau dann, wenn $u = v$ gilt,
2. $\varrho(u, v) \leqq \varrho(u, w) + \varrho(v, w)$ für jedes Elemententripel $u, v, w \in R$ (Dreiecksgleichung).

Daraus folgt dann die Symmetrie, $\varrho(u, v) = \varrho(v, u)$, und die Definitheit, $\varrho(u, v) \geqq 0$, des Abstandes. Man beweist leicht, daß die Axiome (2.1e) und (2.1g) in dem obigen System äquivalent sind den Axiomen:

(2.1e′) Aus $\tau_n \leqq \varrho_n \leqq \sigma_n$ für $n = 1, 2, \ldots$ und $\lim \tau_n = \lim \sigma_n = \sigma$ folgt $\lim \varrho_n = \sigma$

bzw.

(2.1g′) aus $0 \leqq \varrho_n \leqq \tau$ für $n \geqq n_0$ und $\lim \varrho_n = 0$ folgt für jede natürliche Zahl k die Existenz einer Nummer $N(k)$ mit $\varrho_n \leqq \frac{\tau}{k}$ für $n \geqq N(k)$.

Die *Konvergenz* einer Folge $u_n \to u$, $n = 0, 1, 2, \ldots$ in R ist stets im Sinne der sog. *starken Konvergenz* bezüglich der Metrik N zu verstehen, d.h.

$$u = \lim_{n \to \infty} u_n \quad \textit{ist äquivalent} \quad \lim_{n \to \infty} \varrho(u_n, u) = 0. \tag{2.2}$$

* Die Axiome a) bis f) sind wörtlich von J. Schröder [9] übernommen. Zum Beweis des in § 3 aufgestellten Satzes über Iterationsverfahren mit veränderlichen Operatoren (1.3) erwies sich das zusätzliche Axiom g) als erforderlich, dessen Unabhängigkeit von den übrigen etwa durch das Beispiel $\varrho_n = f_n(x) = \frac{n\,x}{1 + n^2\,x^2}$ und $\tau = \frac{1}{2}$ im Raum der stetigen Funktionen gezeigt wird.

Man zeigt leicht:

(2.3) *Aus* $u_n \to u$, $v_n \to v$ *folgt* $\varrho(u_n, v_n) \to \varrho(u, v)$.

Außerdem wird R als *vollständig* vorausgesetzt, d.h. für jede Folge $\{u_n\}$ in R, für die das Cauchysche Konvergenzkriterium $\lim_{l,k\to\infty} \varrho(u_l, u_k) = 0$ gilt, existiert ein Grenzelement $u = \lim_{n\to\infty} u_n$ in R.

Der bezüglich N metrische Raum R heißt *lokalkompakt*, wenn jede unendliche Folge $u_n \in R$, $n = 0, 1, 2, \ldots$, für die $\varrho(u_n, v) \leqq \sigma$, $n = 0, 1, 2, \ldots$, mit $v \in R$ und $\sigma \in N$ gilt, eine konvergente Teilfolge u_{k_n} besitzt.

Eine Folge von Operatoren T_n, $n = 0, 1, 2, \ldots$, die in einem Gebiet D von R definiert sind, konvergiert dort *stetig** gegen einen Operator T, wenn für jede Folge $\{v_n\}$ aus D, die gegen einen Punkt v von D konvergiert, der Grenzwert $\lim_{n\to\infty} T_n v_n$ existiert und gleich Tv ist.

Eine Folge von Operatoren T_n, $n = 0, 1, 2, \ldots$, konvergiert in einem Bereich G ihres Definitionsbereiches *gleichmäßig* gegen T, wenn $\varrho(T_n u, Tu) \leqq \sigma$ für alle $u \in G$ mit einer Größe $\sigma \in N$ ist und nach Vorgabe einer positiven Zahl k eine Nummer $n(k)$ existiert derart, daß

$$\varrho(T_n u, Tu) \leqq \frac{\sigma}{k} \quad \text{für} \quad n \geqq n(k) \text{ und } \textit{alle } u \in G$$

gilt.

Ein in einem Bereich G eines bezüglich N metrischen Raumes R definierter Operator T heißt dort *beschränkt*, wenn

(2.4) $$\varrho(Tu, Tv) \leqq P\varrho(u, v), \qquad u, v \in G,$$

mit einem linearen, stetigen und positiven Operator P gilt. Dabei heißt P *linear*, wenn $P(\alpha\varrho + \beta\sigma) = \alpha P\varrho + \beta P\sigma$ (α, β reelle Zahlen, $\varrho, \sigma \in N$) gilt, *stetig*, wenn $\lim P\varrho_n = P\varrho$ aus $\lim \varrho_n = \varrho$, und *positiv*, wenn $P\varrho \geqq 0$ aus $\varrho \geqq 0$ folgt.

Ein linearer positiver Operator P ist auch monoton, d.h. es gilt

(2.5) *aus* $\varrho \geqq \sigma$, P *linear, positiv, folgt* $P\varrho \geqq P\sigma$.

Schließlich verwenden wir noch folgenden einfachen

Hilfssatz 1: *Für die Folge von Operatoren* T_n, $n = 0, 1, 2, \ldots$ *gelte in einem Bereich* G *ihres gemeinsamen Definitionsgebietes* $D \leqq R$

(2.6) $$\lim_{n\to\infty} \varrho(Tu, T_n u) = 0, \qquad u \in G,$$

und

(2.7) $$\varrho(T_n u, T_n v) \leqq P\varrho(u, v), \qquad n = 0, 1, 2, \ldots, \qquad u, v \in G,$$

mit einem stetigen Operator P. *Dann konvergieren die Operatoren* T_n, $n = 0, 1, 2, \ldots$, *stetig in* G.

Beweis. Es gelte $\lim_{n\to\infty} v_n = v$, $v_n, v \in G$. Dann folgt nach der Dreiecksungleichung und (2.7):

$$0 \leqq \varrho(T_n v_n, Tv) \leqq \varrho(T_n v_n, T_n v) + \varrho(T_n v, Tv) \leqq P\varrho(v_n, v) + \varrho(T_n v, Tv).$$

Hieraus ergibt sich wegen (2.6) und der Stetigkeit von P mit (2.1c) und (2.1e) die Behauptung.

* Vgl. C. CARATHÉODORY [13] (S. 1ff.).

§ 3. Der Fixpunktsatz

Wir treffen folgende *allgemeine Voraussetzungen:*

A. *Es sei R ein bezüglich eines linearen halbgeordneten Raumes N metrischer, vollständiger Raum**. *Dabei sei die Konvergenz in N durch die Axiome* (2.1a—g) *festgelegt.*

B. *Ferner sei eine Folge von in einem gemeinsamen Teilgebiet D von R definierten Operatoren*

$$T_0, T_1, T_2, \ldots \tag{3.1}$$

gegeben mit den Eigenschaften:

1a. *Die Folge* (3.1) *konvergiert gleichmäßig**,** *in D gegen einen Operator T*

$$\lim_{n\to\infty} \varrho(T_n u, T u) = 0, \qquad u \in D. \tag{3.2}$$

Statt der gleichmäßigen Konvergenz können wir fordern:

1a'. *Der Raum R ist lokalkompakt** *und es gilt* (3.2) *ohne die Bedingung* (2.1g).

2. *Die Operatoren* (3.1) *seien in einem beschränkten Teilgebiet G von D gleichmäßig beschränkt, d.h. es existiert ein stetiger, positiver, linearer Operator P, so daß*

$$\varrho(T_n u, T_n v) \leqq P\varrho(u, v), \qquad n = 0, 1, 2, \ldots, \qquad u, v \in G \subseteqq D, \tag{3.3}$$

gilt.

3. *Schließlich existiere eine obere Schranke $\mu \in N$ derart, daß an einer Stelle u_0 von G*

$$\varrho(T_0 u_0, T_n u_0) \leqq \mu, \qquad n = 0, 1, 2, \ldots, \qquad u_0 \in G, \tag{3.4}$$

gilt.

Es ergeben sich einige einfache Folgerungen aus A und B:

Zunächst folgt aus (3.2) und (3.3), daß auch der Grenzoperator T beschränkt ist mit demselben Operator P, d.h. es gilt

$$\varrho(T u, T v) \leqq P\varrho(u, v), \qquad u, v \in G. \tag{3.5}$$

Dies folgt aus (2.3) und den Axiomen (2.1a, c, d, f) auf die Folge

$$\tau_n = P\varrho(u, v) - \varrho(T_n u, T_n v) \geqq 0, \qquad n = 0, 1, 2, \ldots,$$

angewandt.

Aus (3.3) und (3.4) folgt für das gesamte beschränkte Gebiet G die Existenz einer oberen Schranke für $\varrho(T_m u, T_n u)$: Zunächst folgt aus der Dreiecksungleichung (3.3) und (3.4) für $u \in G$

$$\begin{aligned}\varrho(T_0 u, T_n u) &\leqq \varrho(T_0 u, T_0 u_0) + \varrho(T_0 u_0, T_n u_0) + \varrho(T_n u_0, T_n u) \\ &\leqq 2P\varrho(u, u_0) + \mu \leqq \tilde{\mu}, \qquad n = 0, 1, 2, \ldots,\end{aligned}$$

* Bezügl. der Definitionen siehe § 2.

** Später werden wir uns noch von dieser Voraussetzung befreien.

denn da G beschränkt ist, liegt $\varrho(u, u_0)$ und damit auch $2P\varrho(u, u_0)$ unter einer festen Schranke. Weiter folgt aus der letzten Gleichung

$$\varrho(T_m u, T_n u) \leqq \varrho(T_0 u, T_m u) + \varrho(T_0 u, T_n u) \leqq 2\tilde{\mu}, \quad n, m = 0, 1, 2, \ldots, \quad u \in G. \tag{3.6}$$

Läßt man hierin $n \to \infty$ gehen, so ergibt sich nach einfacher Überlegung wegen der Stetigkeit des Abstandes [(2.3)] auch

$$\varrho(T_m u, Tu) \leqq 2\tilde{\mu}, \quad m = 0, 1, 2, \ldots, \quad u \in G. \tag{3.7}$$

Schließlich folgt nach Hilfssatz 1 noch, daß die Folge der Operatoren T_n in G stetig gegen T konvergiert.

Unter den allgemeinen Voraussetzungen A und B gilt der

Fixpunktsatz: *Es gelte**

$$\tau_0 \geqq \varrho(T_n u_0, T_0 u_0), \quad n = 0, 1, 2, \ldots.$$

*Liegen u_0 und alle u, für die***

$$\varrho(u, u_1) \leqq (E - P)^{-1} P \varrho(u_0, u_1) + (E - P)^{-1} \tau_0 = \sigma \tag{3.8}$$

gilt, in dem Gültigkeitsbereich G der Gleichung (3.3) und konvergiert die Neumannsche Reihe

$$\sum_{\nu=0}^{\infty} P^\nu \varrho \quad \text{für alle } \varrho \in N, \tag{3.9}$$

so konvergiert das Iterationsverfahren

$$u_{n+1} = T_n u_n \tag{3.10}$$

gegen die einzige Lösung u der Gleichung

$$u = Tu$$

*in G und es gilt die Fehlerabschätzung****

$$\varrho(u, u_1) \leqq (E - P)^{-1} P \varrho(u_0, u_1) + (E - P)^{-1} \varrho(Tu_0, u_1). \tag{3.11}$$

Beweis. α) Mit P ist** auch $(E - P)^{-1}$ ein positiver Operator. Daher ist die rechte Seite von (3.8) $\sigma \geqq 0$. Es liegt also u_1 in dem durch (3.8) gegebenen Bereich von G. Wir treffen die Induktionsannahme, daß $u_1, u_2, \ldots, u_s$ in G liegen. Dann sei für $n = 1, 2, \ldots, s+1$

$$\tau_n \geqq \varrho(T_l u_n, T_n u_n) \quad \text{für} \quad l \geqq n. \tag{3.12}$$

Die Existenz dieser oberen Schranken τ_n folgt aus (3.4) und (3.6).

---

* Die Existenz einer solchen oberen Schranke ergibt (3.4), z.B. $\tau_0 = \mu$. Existiert eine kleinste obere Schranke, so wird man zweckmäßig $\tau_0 = \sup_n \varrho(T_n u_0, T_0 u_0)$ setzen.

** Die Existenz des Operators $(E - P)^{-1}$ folgt aus der Konvergenz der Reihe (3.9), und es gilt $(E - P)^{-1} \varrho = \sum_{\nu=0}^{\infty} P^\nu \varrho$, $\varrho \in N$, siehe J. SCHRÖDER [9].

*** Diese Fehlerabschätzung läßt sich natürlich auch mit u_{n+1} und u_n statt u_1 und u_0 schreiben, was nur eine andere Zählung bedeutet.

Sodann beweisen wir durch vollständige Induktion

$$\varrho(u_{n+k}, u_n) \leqq (E-P)^{-1} P\, \varrho(u_n, u_{n-1}) + (E-P)^{-1} \tau_{n-1} = \sigma_n \quad \text{für} \quad 1 \leqq n \leqq s \quad \text{und} \quad k \geqq 0. \tag{3.13}$$

Es ist $\sigma_n \geqq 0$ für alle n. Daher ist (3.13) für $k=0$ und $1 \leqq n \leqq s$ erfüllt. (3.13) sei für $k=0, 1, \ldots, r$ erfüllt, und es mögen $u_{n-1}, u_n, u_{n+1}, \ldots, u_{n+r}$ in dem Gültigkeitsbereich G von (3.3) liegen. Dann folgt aus (3.3), (3.12) und der Dreiecksungleichung

$$\begin{aligned} \varrho(u_{n+r+1}, u_n) &= \varrho(T_{n+r} u_{n+r}, T_{n-1} u_{n-1}) \\ &\leqq \varrho(T_{n+r} u_{n+r}, T_{n+r} u_{n-1}) + \varrho(T_{n+r} u_{n-1}, T_{n-1} u_{n-1}) \\ &\leqq P\, \varrho(u_{n+r}, u_{n-1}) + \tau_{n-1}. \end{aligned} \tag{3.14}$$

Da nach (2.5) der lineare, positive Operator P auch monoton ist, gilt nach der Dreiecksungleichung

$$P\, \varrho(u_{n+r}, u_{n-1}) \leqq P\, \varrho(u_{n+r}, u_n) + P\, \varrho(u_n, u_{n-1}).$$

Setzt man dies in (3.14) ein und sodann für $\varrho(u_{n+r}, u_n)$ die rechte Seite von (3.13), was wegen der Induktionsvoraussetzung für k erlaubt ist, so ergibt sich

$$\varrho(u_{n+r+1}, u_n) \leqq [P(E-P)^{-1} + E]\, P\, \varrho(u_n, u_{n-1}) + [P(E-P)^{-1} + E]\, \tau_{n-1}. \tag{3.15}$$

Nun ist aber

$$[P(E-P)^{-1} + E]\, \varrho = (E-P)^{-1} \varrho \quad \text{für alle } \varrho \in N,$$

denn es gilt

$$[P(E-P)^{-1} + E]\, \varrho - (E-P)^{-1} \varrho = [E - (E-P)(E-P)^{-1}]\, \varrho = [E-E]\, \varrho = 0.$$

Daher folgt aus (3.15) die Beziehung (3.13) für $k=r+1$, also unter den obigen Annahmen für alle $k \geqq 0$ und $1 \leqq n \leqq s$. Setzt man hierin $n=1$, so erkennt man, da für u_0 und u_1 alle getroffenen Induktionsvoraussetzungen gültig sind, daß die Iteration (3.10) nicht aus dem durch (3.8) gegebenen Bereich herausführt. Es liegen damit alle u_n in G und daher folgt die Gültigkeit von (3.12) für alle $k \geqq 0$ und $n \geqq 1$.

β) Wir zeigen weiter, daß

$$\lim_{n \to \infty} \varrho(T_{l_n} u_{n-1}, T_{n-1} u_{n-1}) = 0 \tag{3.16}$$

ist, wobei $\{l_n\}$ eine (nicht notwendig monotone) Folge ganzer Zahlen mit $l_n \geqq n$ ist. Dies ist wegen

$$\varrho(T_{l_n} u_{n-1}, T_{n-1} u_{n-1}) \leqq \varrho(T_{l_n} u_{n-1}, T u_{n-1}) + \varrho(T u_{n-1}, T_{n-1} u_{n-1})$$

der Fall, wenn die Größen

$$\varrho_m = \varrho(T_{k_m} u_m, T u_m) \quad \text{mit} \quad k_m \geqq m$$

für $m \to \infty$ gegen Null konvergieren.

Dies ist aber offenbar wegen der gleichmäßigen Konvergenz (Voraussetzung B 1a.) der Operatoren T_n, $n=0, 1, 2, \ldots$, also auch der Operatoren T_{k_m}, $m=0, 1, 2, \ldots, k_m \geqq m$, gegen T der Fall, weil nach α) alle u_m in dem beschränkten Bereich (3.8) von G liegen.

Ist statt B 1a. die Voraussetzung B 1a'. erfüllt, so zeigen wir die Konvergenz der Folge $\varrho_m \to 0$ für $m \to \infty$ folgendermaßen:

Nach (3.7) und $u_m \in G$ für $m = 0, 1, 2, \ldots$ (siehe α)) liegen die Abstände ϱ_m alle unter einer festen Schranke τ. Würde ϱ_m nicht gegen Null konvergieren, so gäbe es eine positive Zahl p und eine Teilfolge ϱ_{r_n}, $n = 1, 2, \ldots$, für die die Gleichung $\varrho_{r_n} \leqq \frac{\tau}{p}$ nicht erfüllt ist. Wir schreiben dafür*

$$\varrho_{r_n} \nleqq \frac{\tau}{p}, \qquad n = 1, 2, \ldots . \tag{3.17}$$

Die unendliche Folge u_{r_n} besitzt wegen der Lokalkompaktheit des Teilraumes G von R eine konvergente Teilfolge. Wir können daher zugleich mit (3.17) annehmen, daß bereits u_{r_n}, $n = 1, 2, \ldots$, gegen ein Element v von G konvergiert. Wegen der stetigen Konvergenz der Operatoren T_n (Hilfssatz 1) konvergiert** damit jede Folge $T_{l_n} u_{r_n}$, $n = 1, 2, \ldots, l_n \geqq r_n$, gegen Tv, und hieraus folgt mittels der Dreiecksungleichung

$$\lim_{n \to \infty} \varrho_{r_n} = 0 .$$

Dieses steht aber nach dem Axiom (2.1g') im Widerspruch zu (3.17). Hieraus folgt, daß die Größen ϱ_m gegen Null gehen mit $m \to \infty$, und damit ist auch die Behauptung (3.16) bewiesen.

γ) Wir wollen jetzt zeigen, daß die Folge $\{u_n\}$ gegen ein Element u konvergiert. Zu diesem Zweck beweisen wir zunächst die Ungleichung

$$\varrho(u_{n+r+k}, u_{n+k-1}) \leqq P^k \varrho(u_{n+r}, u_{n-1}) + \sum_{\nu=1}^{k} P^{\nu-1} \tau_{n+k-\nu-1}, \qquad n, k \geqq 1, \tag{3.18}$$

wobei die τ_i als obere Schranken gemäß (3.12) definiert sind und $r = r(n, k)$ eine beliebige Doppelfolge natürlicher Zahlen bedeute.

Für $k = 1$ erhalten wir nach (3.3), (3.12) und der Dreiecksungleichung

$$\begin{aligned} \varrho(u_{n+r+1}, u_n) &= \varrho(T_{n+r} u_{n+r}, T_{n-1} u_{n-1}) \\ &\leqq \varrho(T_{n+r} u_{n+r}, T_{n+r} u_{n-1}) + \varrho(T_{n+r} u_{n-1}, T_{n-1} u_{n-1}) \\ &\leqq P \varrho(u_{n+r}, u_{n-1}) + \tau_{n-1}, \qquad n = 1, 2, \ldots . \end{aligned} \tag{3.19}$$

Also ist (3.18) für $k = 1$ und alle n und r erfüllt. Es sei (3.18) für $k = 1, 2, \ldots, s$ erfüllt. Ersetzt man in (3.19) n durch $n + s$, so ergibt sich unter Benutzung von (3.18) für $k = s$

$$\begin{aligned} \varrho(u_{n+r+s+1}, u_{n+s}) &\leqq P \varrho(u_{n+s+r}, u_{n+s-1}) + \tau_{n+s-1} \\ &\leqq P^{s+1} \varrho(u_{n+r}, u_{n-1}) + P \sum_{\nu=1}^{s} P^{\nu-1} \tau_{n+s-\nu-1} + \tau_{n+s-1} \\ &\leqq P^{s+1} \varrho(u_{n+r}, u_{n-1}) + \sum_{\nu=1}^{s+1} P^{\nu-1} \tau_{n+(s+1)-\nu-1} . \end{aligned}$$

Also ist (3.18) auch für $k = s + 1$ und daher für alle $k \geqq 1$ erfüllt.

Nach α) und (3.8) gilt

$$\varrho(u_m, u_n) \leqq \varrho(u_m, u_1) + \varrho(u_n, u_1) \leqq 2\sigma, \qquad m, n = 1, 2, \ldots .$$

* Man beachte, daß aus $\varrho \nleqq \sigma$ noch nicht $\varrho \geqq \sigma$ folgt.

** Mit T_n konvergiert auch die Folge T_{l_n}, $n = 1, 2, \ldots, l_n \geqq n$, stetig, wie man leicht unter Verwendung des Axioms (2.1b) schließt.

Daher ist auch $\varrho(u_{n+r}, u_{n-1}) \leqq 2\sigma$ für $n \geqq 2$, $r \geqq 0$, und da P monoton und linear ist,

$$P^k \varrho(u_{n+r}, u_{n-1}) \leqq 2P^k\sigma, \qquad \sigma \in N. \tag{3.20}$$

Wegen der Konvergenz der Neumannschen Reihe für alle $\varrho \in N$ folgt, wie man leicht zeigen kann,

$$2P^k\sigma \to 0 \quad \text{mit} \quad k \to \infty,$$

und daher geht nach (2.1e) auch die linke Seite von (3.20), also der erste Summand der rechten Seite von (3.18), bei beliebigen $n \geqq 1$ und $r \geqq 0$ mit $k \to \infty$ gegen Null.

Nach β) konvergiert die rechte Seite von (3.12) gegen Null. Wir können daher wegen (2.1g') auch immer erreichen, daß die τ_n eine Nullfolge bilden*. Es seien nun t_n obere Schranken der τ_m für $m \geqq n-1$. Dann können wir wegen $\tau_m \to 0$ für $m \to \infty$ wiederum nach (2.1g') annehmen, daß auch die $t_n \to 0$ gehen mit $n \to \infty$. Mit

$$t_n \geqq \tau_{n+k-\nu-1}, \qquad t_n \to 0 \quad \text{mit} \quad n \to \infty$$

ergibt sich für den zweiten Summanden der rechten Seite von (3.18)

$$\sum_{\nu=1}^{k} P^{\nu-1} \tau_{n+k-\nu-1} \leqq \sum_{\nu=1}^{k} P^{\nu-1} t_n \leqq \sum_{\nu=1}^{\infty} P^{\nu-1} t_n = (E-P)^{-1} t_n \to 0 \quad \text{mit} \quad n \to \infty.$$

Nach (2.1e) geht daher der zweite Summand von (3.18) mit $n \to \infty$ für alle k gegen Null. Hieraus folgt zusammen mit den obigen Betrachtungen und (2.1c, e)

$$\lim_{n,k \to \infty} \varrho(u_{n+r+k}, u_{n+k-1}) = 0.$$

Da $r = r(n, k)$ willkürliche ganze Zahlen $\geqq 0$ sind, ist das Cauchysche Konvergenzkriterium erfüllt. Daher existiert wegen der Vollständigkeit von R ein Grenzelement

$$u = \lim_{n \to \infty} u_n. \tag{3.21}$$

Für dieses Element u folgt wegen $\varrho(u_n, u_1) \leqq \sigma$ (nach α) und (2.3), (2.1a, c, d, f) auf $\sigma - \varrho(u_n, u_1)$ angewandt) auch

$$\varrho(u, u_1) \leqq \sigma. \tag{3.22}$$

Damit haben wir zunächst die Fehlerabschätzung (3.8). Es gilt jedoch auch die in vielen Fällen schärfere Abschätzung (3.11), aus der wegen $\tau_0 \geqq \varrho(Tu_0, T_0u_0)$ die Abschätzung (3.8) folgt; denn führen wir nur einen Schritt** $u_1 = T_0u_0$ der Iteration aus, so hängt der Fehler $\varrho(u, u_1)$ offenbar nicht von den weiteren Operatoren T_n, $n \geqq 1$, sondern nur noch von dem Operator T ab. Wir können daher für die Fehlerabschätzung $T_n = T$, $n = 1, 2, \ldots$, annehmen und erhalten so als kleinste obere Schranke statt τ_0 die Größe $\varrho(Tu_0, u_1)$ und damit (3.11).

* Unter der Annahme der Existenz einer kleinsten oberen Schranke für die obigen beschränkten Folgen in N wäre der Beweis ein wenig einfacher mit dem Supremum. Wir setzen dies jedoch nicht voraus.

** Vgl. die 3. Anmerkung, S. 50.

δ) $u = \lim_{n\to\infty} u_n$ ist Lösung von (1.1). Denn es folgt aus

$$\begin{aligned} 0 \leqq \varrho(u, Tu) &\leqq \varrho(u, T_n u_n) + \varrho(T_n u_n, Tu) \\ &\leqq \varrho(u, u_{n+1}) + \varrho(T_n u_n, T_n u) + \varrho(T_n u, Tu) \\ &\leqq \varrho(u, u_{n+1}) + P\,\varrho(u_n, u) + \varrho(T_n u, Tu) \end{aligned}$$

und anschließendem Grenzübergang $n\to\infty$ nach (3.21) und (3.2), der Stetigkeit von P sowie (2.1 b, c, e) $\varrho(u, Tu) = 0$, also $u = Tu$.

ε) Gäbe es zwei Lösungen u und $v \neq u$ in G, so wäre nach (3.5)

$$0 \leqq \varrho(u, v) = \varrho(Tu, Tv) \leqq P\,\varrho(u, v) \leqq \cdots \leqq P^n \varrho(u, v),$$

und hieraus folgt wegen der Konvergenz der Reihe (3.9) $\lim_{n\to\infty} P^n \varrho = 0$, also nach (2.1 a, e) $\varrho(u, v) = 0$ oder $u = v$ im Widerspruch zur Annahme, w.z.b.w.

Wir bringen noch folgenden *Zusatz zum Fixpunktsatz:*

Wird die Existenz einer Lösung u der Gleichung $u = Tu$ in G vorausgesetzt, so kann die Forderung (3.2) *der gleichmäßigen Konvergenz der Folge $\{T_n\}$ gegen T für alle u von G wesentlich abgeschwächt werden. Der Satz gilt bereits mit Ausnahme der Eindeutigkeitsaussage, wenn man statt* (3.2) *nur voraussetzt, daß die Folge $\{T_n\}$ in dem Punkte $u = Tu$ der Lösung in G konvergiert:*

$$\lim_{n\to\infty} \varrho(Tu, T_n u) = 0 \quad \textit{mit} \quad u = Tu \in G.$$

Beweis. Der Teil α) des Beweises kann genau so geführt werden. Es liegen also alle u_n in G. Da auch u nach Voraussetzung in G liegt, gilt

$$\begin{aligned} \varrho(u_{n+1}, u) = \varrho(T_n u_n, Tu) &\leqq \varrho(T_n u_n, T_n u) + \varrho(T_n u, Tu) \\ &\leqq P\,\varrho(u_n, u) + \varrho(T_n u, Tu). \end{aligned}$$

Durch mehrmalige Anwendung dieser Abschätzung ergibt sich

$$\varrho(u_{n+k}, u) \leqq P^k \varrho(u_n, u) + \sum_{\nu=1}^{k} P^{\nu-1} \varrho(T_{n+k-\nu} u, Tu). \tag{3.23}$$

Da alle u_n in dem durch (3.8) gegebenen Bereich liegen, folgt

$$\varrho(u_n, u) \leqq \varrho(u_n, u_1) + \varrho(u_1, u) \leqq \sigma + \varrho(u_1, u) = \tau, \qquad \tau \in N.$$

Hieraus folgt wegen der Konvergenz der Neumannschen Reihe (3.9) für alle Punkte von N und nach (2.1 e)

$$\lim_{k\to\infty} P^k \varrho(u_n, u) = 0. \tag{3.24}$$

Nach (3.7) liegen die Größen $\varrho_n = \varrho(T_{n+k-\nu} u, Tu)$ unter einer gemeinsamen oberen Schranke: $\varrho_n \leqq 2\tilde{\mu}$.

Da nach Voraussetzung und (2.1 b) $\lim_{n\to\infty} \varrho_n = 0$ ist, existiert wegen (2.1 g' nach Vorgabe einer positiven Zahl m eine Zahl $n_0(m)$, so daß

$$\varrho_n \leqq \frac{2\tilde{\mu}}{m} \quad \text{für} \quad n \geqq n_0(m) \quad \text{und} \quad k \geqq \nu$$

gilt. Hieraus folgt wegen der Linearität und Positivität des Operators P die Abschätzung

$$\sum_{\nu=1}^{k} P^{\nu-1}\,\varrho(T_{n+k-\nu}\,u,\,T\,u) \leqq \sum_{\nu=1}^{k} P^{\nu-1}\,\frac{2\tilde{\mu}}{m} \leqq \frac{2}{m}\sum_{\nu=1}^{\infty} P^{\nu-1}\tilde{\mu}, \qquad n \geqq n_0(m).$$

Hieraus schließt man leicht, daß der 2. Summand auf der rechten Seite von (3.23) mit $n\to\infty$ gegen Null geht. Zusammen mit (3.24) ergibt sich sodann die Behauptung

$$\lim_{s\to\infty} \varrho(u_s, u) = 0.$$

Die Fehlerabschätzung folgt wie oben.

Ferner ist folgende Bemerkung für die Anwendungen häufig nützlich:

Die Voraussetzung (3.2) ist in vielen Fällen für die Anwendung des Satzes störend, z.B. in den Fällen, wo der Operator T nicht explizit bekannt ist, sondern nur ein Näherungsoperator $\widetilde{T}$. Setzen wir $T_n=\widetilde{T}$ für alle n, so ist (3.2) nicht erfüllt. In diesem Falle können wir natürlich auch keine Aussage darüber machen, ob die Folge u_n, $n=0, 1, 2, \ldots$ gegen eine Lösung $u=Tu$ konvergiert. Das wird i. allg. nicht der Fall sein.

Trotzdem gilt die Fehlerabschätzung (3.11) auch dann noch ohne die Forderung (3.2), wenn wir statt dessen (3.5) fordern und wenn (3.8) mit $\tau_0\geqq\varrho(T_n u_0, T_0 u_0)$ und $\tau_0\geqq\varrho(T_n u_0, T u_0)$ erfüllt ist. Beides läßt sich in vielen Fällen zeigen, auch wenn T nicht genau bekannt ist.

Diese Folgerungen ergeben sich aus der einfachen Überlegung, daß wir, wenn wir nur n Schritte durchführen, die Operatoren T_r für $r\geqq n$ uns durch T ersetzt denken dürfen, da wir sie ja gar nicht mehr benutzen. Dabei bleibt offenbar der Beweis des Satzes unverändert. Somit gilt der

Zusatz 2: *Es gelten die Voraussetzungen A und B außer* (3.2). *Ferner gelte*

$$\varrho(Tu, Tv) \leqq P\,\varrho(u, v), \qquad u, v\in G,$$

mit einem linearen, stetigen und positiven Operator P.

Liegen dann u_0 und alle u, für die

$$\varrho(u, u_1) \leqq (E-P)^{-1} P\,\varrho(u_0, u_1) + (E-P)^{-1}\tau_0$$

mit einer Größe $\tau_0\geqq\varrho(T_n u_0, T_0 u_0)$ und $\tau_0\geqq\varrho(Tu_0, T_0 u_0)$ gilt, in G und konvergiert die Reihe $\sum_{\nu=0}^{\infty} P^\nu\varrho$ für alle $\varrho\in N$, so existiert genau eine Lösung u der Gleichung $u=Tu$ in G und es gilt mit $u_{n+1}=T_n u_n$, $n=0, 1, \ldots, s$, die Fehlerabschätzung

$$\varrho(u, u_{s+1}) \leqq (E-P)^{-1} P\,\varrho(u_s, u_{s+1}) + (E-P)^{-1}\varrho(Tu_s, u_{s+1}) \tag{3.25}$$

für jedes $s\geqq 0$.

Wählen wir hierin speziell $T_n=E$, $n=0, 1, 2, \ldots$, so folgt wegen $u_n=u_0$ für alle n, da (3.4) erfüllt ist und die Bedingung (3.3) im Beweis des Fixpunktsatzes nur für Größen u, v aus der Menge der u_n verwendet wird, der

Zusatz 3: *Gilt die allgemeine Voraussetzung A und*

$$\varrho(Tu, Tv) \leqq P\,\varrho(u, v), \qquad u, v\in G,$$

mit einem linearen, stetigen und positiven Operator P, liegen ferner u_0 und alle u, für die

$$\varrho(u, u_0) \leqq (E - P)^{-1} \varrho(T u_0, u_0) = \sigma \tag{3.26}$$

gilt, in G und konvergiert schließlich die Reihe $\sum_{\nu=0}^{\infty} P^\nu \varrho$ für alle $\varrho \in N$, so existiert genau eine Lösung u der Gleichung $u = Tu$ in G, und für diese gilt die Abschätzung (3.26).

Dieser auch leicht auf andere Weise herleitbare Spezialfall ist ein in vielen Fällen sehr einfach anzuwendender Existenzsatz mit Fehlerabschätzung und Eindeutigkeitsaussage.

§ 4. Beispiele

1. Gewöhnliche nichtlineare Gleichungen. Nach unserem Fixpunktsatz ist es möglich, eine Begründung und Fehlerabschätzung für das bekannte Verfahren zu geben, daß man bei der Iteration

$$x_{n+1} = f(x_n) \tag{4.1}$$

die rechte Seite in eine Potenzreihe nach Potenzen von Funktionen $\varphi_j(x)$, $j = 1, 2, \ldots, s$, entwickelt und jeweils nur Glieder bis zu einer festen Ordnung berücksichtigt.

Sei etwa in einem Konvergenzgebiet $x \in B$, $|\varphi_j(x)| \leqq c_j$, die Funktion $f(x)$ in eine Reihe der Form

$$f(x) = \sum_{\alpha_1+\alpha_2+\cdots+\alpha_s=0}^{\infty} a_{\alpha_1\alpha_2\ldots\alpha_s}(x)\,[\varphi_1(x)]^{\alpha_1}[\varphi_2(x)]^{\alpha_2}\ldots[\varphi_s(x)]^{\alpha_s} \tag{4.2}$$

entwickelbar und werden beim n-ten Schritt nur Potenzen bis zur Ordnung $\alpha_1+\alpha_2+\cdots+\alpha_s \leqq r_n$ berücksichtigt, so hat man statt (4.1) das Verfahren

$$\begin{aligned} x_{n+1} &= \sum_{\alpha_1+\alpha_2+\cdots+\alpha_s\leqq r_n} a_{\alpha_1\ldots\alpha_s}(x_n)\,[\varphi_1(x_n)]^{\alpha_1}\ldots[\varphi_s(x_n)]^{\alpha_s} \\ &= f(x_n) - \sum_{\alpha_1+\alpha_2+\cdots+\alpha_s\geq r_n+1}^{\infty} a_{\alpha_1\ldots\alpha_s}(x_n)\,[\varphi_1(x_n)]^{\alpha_1}\ldots[\varphi_s(x_n)]^{\alpha_s} \\ &= f_n - \Sigma_n = T_n x_n. \end{aligned} \tag{4.3}$$

Mit gewöhnlichem Zahlenabstand ($|x-y|$) hat man dann mit $Tx = f(x)$ in der Fehlerabschätzung (3.11) bzw. (3.25) für $\varrho(T x_n, x_{n+1})$ die Größe $|\Sigma_n|$ einzusetzen.

1. Beispiel: Gegeben ist die Gleichung*

$$x^3 = 3R^3\left(1 + \frac{r}{x} + \frac{4r^2}{3x^2} + \frac{5r^3}{3x^3} + \frac{r^3}{R^3}\right). \tag{4.4}$$

Es handelt sich um ein Problem bei der Bewegung eines Satelliten um einen Hauptplaneten. Dabei ist r der Radius des Satelliten, R im wesentlichen der des Hauptplaneten und x der Abstand beider Mittelpunkte. Hieraus ergibt sich, daß x, r und R reelle positive Zahlen sind. Daher gilt für die gesuchte Lösung ξ

* D. VAUGHAN [14].

von (4.4): $\xi \geqq \sqrt[3]{3}R$. Dabei bedeutet ξ einen kritischen Abstand, unter dem der Satellit nicht mehr nur durch innere Gravitationskräfte zusammenhalten würde.

Die Lösung wurde in [14]* näherungsweise bestimmt, indem x/R nach Potenzen von r/R unter Vernachlässigung der 4. und höheren Potenzen von r/R entwickelt wurde.

Setzen wir zur Abkürzung $\sqrt[3]{3}=\alpha$, so haben wir

$$x = \alpha R \sqrt[3]{1 + \frac{r}{x} + \frac{4r^2}{3x^2} + \frac{5r^3}{3x^3} + \frac{r^3}{R^3}} = f(x).$$

Mit den obigen Bezeichnungen können wir schreiben: $\varphi_1(x)=r/x$ und $\varphi_2(x)=r/R$. Damit haben wir das Iterationsverfahren (4.3). Wählen wir etwa $x_0=\alpha R$, $r_0=2$, $r_1=2$, $r_n=3$ für $n\geqq 2$, so ergibt sich nach leichter Rechnung

$$x_1 = \alpha R\left(1+\frac{r}{3\alpha R}+\frac{r^2}{3\alpha^2R^2}\right); \quad x_2 = \alpha R\left(1+\frac{r}{3\alpha R}+\frac{2r^2}{9\alpha^2R^2}\right)$$

und

$$x_3 = \alpha R\left(1+\frac{r}{3\alpha R}+\frac{2r^2}{9\alpha^2R^2}+\frac{32r^3}{81\alpha^3R^3}\right).$$

Das Verfahren kommt zum Stillstand, $x_n=x_3$ für $n\geqq 3$, wenn man sich auf die Entwicklung der Lösung bis zur 3. Potenz von r/R beschränkt. Daher bekommt die Fehlerabschätzung (3.11) bzw. (3.25), wenn wir als Abstand $\varrho(x,\xi)$ den Betrag der Differenz wählen, die einfache Gestalt

$$|\xi - x_3| \leqq \frac{1}{1-P}\,|T_3 x_3 - f(x_3)|,$$

wobei hier P eine positive Zahl ist, z.B. das Maximum der Beträge von $f'(x)$ und $T_3'(x)$ in dem Bereich G.

Wir wählen für G den Bereich $x\geqq\sqrt[3]{3}R$ und zur Illustration** $r/R=0{,}01$. Dann erhält man nach leichter Rechnung:

$$x_3 = T_3(x_3) = 1{,}445\,598\,49R \quad \text{und} \quad f(x_3) = 1{,}445\,598\,50R, \quad \text{sowie } P = 0{,}0024$$

und damit das Ergebnis:

$$|\xi - x_3| < \frac{1}{1-0{,}0024}\cdot 2\cdot 10^{-8} < 2{,}005\cdot 10^{-8}$$

oder

$$\underline{\xi = 1{,}445\,598\,50 \pm 2{,}1\cdot 10^{-8}.}$$

Offenbar sind alle Voraussetzungen von Zusatz 2 (§3) erfüllt.

Ein weiteres ganz anders geartetes Beispiel ist das folgende

2. Beispiel: Es ist die reelle positive Lösung ξ der Gleichung

$$z = 10\cdot\zeta(z) \tag{4.5}$$

zu berechnen, wobei $\zeta(z)$ die Riemannsche Zetafunktion bedeutet:

$$\zeta(z) = \sum_{n=1}^{\infty}\frac{1}{n^z}. \tag{4.6}$$

* Dort noch ohne Konvergenznachweis und Fehlerabschätzung.

** Für kleinere Werte von r/R würde man noch eine größere Genauigkeit erhalten.

Hier ist der Operator $Tz = 10 \cdot \zeta(z)$ i. allg. nur näherungsweise bekannt je nach der Anzahl der Glieder der Reihe (4.6), die berücksichtigt werden. Diese Tatsache wird durch die gewöhnliche Iteration $z_{n+1} = Tz_n$ und die bekannten Fehlerabschätzungen nicht mehr erfaßt.

Die Reihe (4.6) konvergiert bekanntlich in dem Bereich Realteil $z = x \geqq 1 + \varepsilon$, $\varepsilon > 0$, gleichmäßig und stellt dort eine stetige Funktion dar. Daher liegt wegen $10 - 10 \cdot \zeta(10) = -0{,}00995 < 0$ und $11 - 10 \cdot \zeta(11) = 0{,}99506 > 0$ eine reelle Lösung von (4.5) in dem Bereich G: $10 \leqq x \leqq 11$.

Wir führen die Iteration

$$x_{k+1} = 10 \cdot \sum_{n=1}^{r_k} \frac{1}{n^{x_k}} = T_k x_k, \qquad k = 0, 1, 2, \ldots,$$

durch, wobei r_k eine nichtfallende Folge natürlicher Zahlen sei. Wählen wir etwa $r_0 = 2$, $r_1 = 3$ und $r_2 = r_3 = 4$, so erhalten wir mit $x_0 = 10$ die Iterationsfolge:

$$x_1 = 10{,}00976563$$
$$x_2 = 10{,}00986718$$
$$x_3 = 10{,}00987600$$
$$x_4 = 10{,}00987593.$$

Als Abstand $\varrho(x, y)$ wählen wir den gewöhnlichen Betrag. Dann erhält man etwa mittels der Abschätzungen

$$\frac{d}{dx} T_k x \Big| < 10\,|\zeta'(x)| = 10 \sum_{n=1}^{\infty} \frac{\log n}{n^x} \leqq 10 \left\{ \sum_{n=2}^{4} \frac{\log n}{n^x} + \int_4^{\infty} \frac{\log t}{t^x}\, dt \right\} < 7 \cdot 10^{-3} = P \quad \text{für} \quad x \geqq 10$$

und

$$|T_4 x_4 - T x_4| = 10 \sum_{n=5}^{\infty} \frac{1}{n^{x_4}} \leqq 10 \left\{ \frac{1}{5^{x_4}} + \int_5^{\infty} \frac{dt}{t^{x_4}} \right\} < 1{,}57 \cdot 10^{-6} \quad \text{für} \quad x \geqq 10,$$

sowie $|x_3 - x_4| < 1 \cdot 10^{-7}$ nach (3.11) die Fehlerschranke

$$|x_4 - \xi| < 1{,}6 \cdot 10^{-6}.$$

2. Systeme nichtlinearer Gleichungen. Eine weitere Anwendung des Fixpunktsatzes bieten die in der praktischen Mathematik häufig auftretenden Fälle, bei denen Gleichungen nur bis auf eine angebbare Genauigkeit ihrer Parameter vorliegen. In diesen Fällen ist der Operator T nicht (genau) bekannt. Wir wenden dann den Zusatz 2 an:

Es sei etwa das nichtlineare Gleichungssystem in impliziter Gestalt

$$\widetilde{F}_i(x^1, x^2, \ldots, x^n) = 0 \qquad (i = 1, 2, \ldots, n) \tag{4.7}$$

gegeben, wobei $\widetilde{F}_i$ andeuten soll, daß die Gleichungen nur näherungsweise vorliegen. Wir schreiben zur Abkürzung für (4.7)

$$\widetilde{G}u = 0 \quad \text{mit dem Vektor} \quad u = (x^1, x^2, \ldots, x^n) \tag{4.8}$$

und verwenden zur Lösung z.B. das vereinfachte Newtonsche Verfahren*

$$u_{n+1} = u_n - (\tilde{G}'_{(u_0)})^{-1}\tilde{G}u_n = \tilde{T}u_n, \tag{4.9}$$

wobei $\tilde{G}'_{(u_0)}$ die an der Stelle u_0 genommene Funktionaldeterminante von (4.7) nach den x^j ist und $(\tilde{G}'_{(u_0)})^{-1}$ deren Kehrmatrix bedeutet, deren Existenz im Falle $\tilde{G}'_{(u)} \neq 0$ in einer Umgebung* $\mathfrak{B}$ der Lösung u von (4.8) gesichert ist.

Wir verwenden als Abstand $||u-v||$ den Vektor, dessen Komponenten die absoluten Beträge von $u-v$ sind, und bezeichnen allgemein mit $||A||$ die Matrix, dessen Elemente die absoluten Beträge der Matrix A sind.

Die Ungenauigkeit der Ausgangsgleichung (4.8) sei z.B. durch

$$||\tilde{G}u_0 - Gu_0|| \leqq \mu_0 \qquad u_0 \in \mathfrak{B} \tag{4.10}$$

beschränkt.

Gilt dann

$$||\tilde{T}u - \tilde{T}v|| \leqq P||u-v|| \quad \text{und} \quad ||Tu - Tv|| \leqq P||u-v|| \quad \text{in } \mathfrak{B}$$

mit einer positiven Matrix P und dem Operator

$$Tu = u - (\tilde{G}'_{(u_0)})^{-1}Gu$$

und kann man außerdem zeigen, daß $(E-P)^{-1}$ existiert und der Bereich

$$||u-u_1|| \leqq (E-P)^{-1}[P||u_0-u_1|| + \tau_0] \tag{4.11}$$

mit $\tau_0 = ||(\tilde{G}'_{(u_0)})^{-1}||\,\mu_0$ in $\mathfrak{B}$ liegt, so folgt die Fehlerabschätzung (3.25) mit $\varrho(u, u_{s+1}) = ||u-u_{s+1}||$. Speziell schätzen wir ab

$$\begin{aligned} ||\tilde{T}u - \tilde{T}v|| &\leqq ||(\tilde{G}'_{(u_0)})^{-1}[\tilde{G}'_{(u_0)}(u-v) - (\tilde{G}u - \tilde{G}v)]|| \\ &\leqq ||(\tilde{G}'_{(u_0)})^{-1}|| \max_{u\in\mathfrak{B}} ||\tilde{G}'_{(u_0)} - \tilde{G}'_{(u)}||\,||u-v|| \end{aligned}$$

und entsprechend

$$||Tu - Tv|| \leqq ||(\tilde{G}'_{(u_0)})^{-1}|| \max_{u\in\mathfrak{B}} ||\tilde{G}'_{(u_0)} - G'_{(u)}||\,||u-v||$$

und wählen die positive Matrix P so, daß $P||u-v||$ größer oder gleich den rechten Seiten dieser Gleichungen wird.

3. Beispiel: Als spezielles Beispiel soll die nichtlineare Behandlung der Matrizeneigenwertaufgabe**,***

$$Ax = \lambda Bx$$

mit den Matrizen

$$A = \begin{pmatrix} 0{,}89689 & -2{,}01310 \\ 1{,}01900 & 1{,}49791 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 0{,}03082 & -1{,}20201 \\ 1{,}13810 & 2{,}39792 \end{pmatrix},$$

---

* Wir setzen hier alle erforderlichen Stetigkeits- und Differenzierbarkeitseigenschaften der $\tilde{F}_i$ voraus.

** Die nichtlineare Behandlung von Matrizeneigenwertaufgaben führte m. W. H. Unger [15] zuerst ein. Vgl. ferner L. Collatz [8], wo auch Verfahren höherer Ordnung zur Lösung benutzt werden.

*** Ohne Berücksichtigung der Ungenauigkeit der Ausgangsmatrizen kann man hier natürlich die Lösungen exakt angeben. Daher könnte der erste Teil der Rechnung fortgelassen werden. Er soll hier nur zur Illustration gebracht werden. Bei Matrizen mit mehr Zeilen würde man hier außerdem wegen der einfacheren Rechnung zweckmäßig im Banachraum mit Zahlenabstand rechnen.

deren Elemente mit einer Genauigkeit von $\pm 5 \cdot 10^{-6}$ gegeben sind, durchgerechnet werden.

Zur Normierung setzen wir die erste Komponente $x^{(1)}$ des Eigenvektors gleich 1. Dann ist die Unbekannte u der Vektor $u = \begin{pmatrix} x^{(2)} \\ \lambda \end{pmatrix}$. Mit $\tilde{G} u = A x - \lambda B x$ folgt

$$\tilde{G}'_{(u_0)} = \begin{pmatrix} -2{,}01310 + 1{,}20201\,\lambda & -0{,}03082 + 1{,}20201\,x^{(2)} \\ 1{,}49791 - 2{,}39792\,\lambda & -1{,}13810 - 2{,}39792\,x^{(2)} \end{pmatrix}. \tag{4.12}$$

Als Ausgangsvektor sei die bereits recht gute Näherung $u_0 = \begin{pmatrix} 0{,}768 \\ 0{,}728 \end{pmatrix}$ gegeben. Dann wird

$$(\tilde{G}'_{(u_0)})^{-1} = \begin{pmatrix} -0{,}824920876 & -0{,}247036883 \\ 0{,}068595906 & -0{,}315061731 \end{pmatrix} \quad \text{und} \quad \tilde{G} u_0 = \begin{pmatrix} 0{,}000440839 \\ 0{,}000171416 \end{pmatrix}.$$

Hieraus ergibt sich für den Zuwachs δ und die neue Näherung u_1:

$$\delta = \begin{pmatrix} 0{,}000406003 \\ 0{,}000023767 \end{pmatrix}; \qquad u_1 = \begin{pmatrix} 0{,}768406003 \\ 0{,}728023767 \end{pmatrix}.$$

Wir wählen nun für $\mathfrak{B}$ den Bereich

$$\|u - u_1\| \leqq 10^{-4} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

In $\mathfrak{B}$ gilt nach (4.12)

$$\|\tilde{G}'_{(u_0)} - \tilde{G}'_{(u)}\| \leqq \|\tilde{G}'_{(u_0)} - \tilde{G}'_{(u_1)}\| + \begin{pmatrix} 1{,}20201 & 1{,}20201 \\ 2{,}39792 & 2{,}39792 \end{pmatrix} \cdot 10^{-4}$$

$$\leqq \begin{pmatrix} 1{,}4870 & 6{,}1822 \\ 2{,}9683 & 12{,}1343 \end{pmatrix} \cdot 10^{-4}.$$

Wegen der Ungenauigkeit der Ausgangsmatrizen sind die Elemente von $\|\tilde{G}'_{(u_0)} - G'_{(u)}\|$ in $\mathfrak{B}$ nach (4.12) um weniger als $1 \cdot 10^{-5}$ größer. Durch Multiplikation mit $\|(\tilde{G}'_{(u_0)})^{-1}\|$ ergibt sich so für den Abschätzungsoperator*

$$P = \begin{pmatrix} 2{,}0672 & 8{,}2047 \\ 1{,}0756 & 4{,}2855 \end{pmatrix} \cdot 10^{-4}.$$

Nach dem Zeilensummenkriterium konvergiert die Reihe $\sum_{\nu=0}^{\infty} P^\nu \varrho$ für alle $\varrho \in N$. Wir können hier jedoch leicht $(E - P)^{-1}$ genau berechnen. Man findet*

$$(E - P)^{-1} = \begin{pmatrix} 1{,}00021 & 0{,}00083 \\ 0{,}00011 & 1{,}00043 \end{pmatrix}.$$

Es ist $T_n = \tilde{T}$ für $n = 0, 1, 2, \ldots$. Daher brauchen wir nur noch $\|\tilde{T} u_0 - T u_0\|$ nach oben abzuschätzen, um die Fehlerabschätzung (3.25) anwenden zu können. Man findet

$$\|\tilde{G} u_0 - G u_0\| \leqq 5 \cdot 10^{-6} (1 + \lambda_0) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ x_0^{(2)} \end{pmatrix} \leqq \begin{pmatrix} 1{,}528 \\ 1{,}528 \end{pmatrix} \cdot 10^{-5}$$

und damit*

$$\|\tilde{T} u_0 - T u_0\| \leqq \|(\tilde{G}'_{(u_0)})^{-1}\| \, \|\tilde{G} u_0 - G u_0\| \leqq \begin{pmatrix} 1{,}6380 \\ 0{,}5863 \end{pmatrix} \cdot 10^{-5},$$

* Dabei wurde stets nach oben abgerundet.

und schließlich ergibt (4.11) die Fehlerabschätzung

$$\|u - u_1\| \leqq \begin{pmatrix}0{,}104\\0{,}055\end{pmatrix} \cdot 10^{-6} + \begin{pmatrix}1{,}6389\\0{,}5898\end{pmatrix} \cdot 10^{-5} = \begin{pmatrix}1{,}6493\\0{,}5923\end{pmatrix} \cdot 10^{-5}.$$

Dabei ist der erste Summand der Summe eine Abschätzung des Fehlers, wenn die Matrizen A und B genau gegeben sind. In diesem Falle hätte man den Bereich $\mathfrak{B}$ jedoch noch kleiner wählen können und würde einen noch kleineren Fehler $\|u - u_1\|$ erhalten. Durch die Ungenauigkeit der gegebenen Matrizen wird der Fehlerbereich wesentlich vergrößert. Offenbar liegt u in $\mathfrak{B}$. Daher ist die Abschätzung gültig.

3. Nichtlineare Randwertaufgaben. Bei nichtlinearen Randwertaufgaben bei gewöhnlichen Differentialgleichungen hat man in vielen Fällen mit Hilfe der Greenschen Funktion, falls sie existiert, eine Möglichkeit zu einer einfachen Fehlerabschätzung zu gelangen*. Dies ist häufig auch bei der Iteration mit veränderlichen Operatoren bzw. Näherungsoperatoren der zweckmäßige Weg, um die Fehlerabschätzung (3.11) durchzuführen.

Es sei etwa die spezielle Randwertaufgabe

$$L[y] = \tilde{f}(x, y), \quad a \leqq x \leqq b, \quad U_i[y] = c_i \quad (i = 1, 2, \ldots, m) \tag{4.13}$$

gegeben, wobei L ein linearer Differentialoperator ist und $\tilde{f}$ wieder andeuten soll, daß die Gleichung nur bis auf eine angebbare Ungenauigkeit vorliegt. Es existiere die Greensche Funktion $G(x, \xi)$ des Randwertproblems

$$L[y] = r(x), \qquad U_i[y] = c_i \qquad (i = 1, 2, \ldots, m).$$

Dann ist (4.13) gleichwertig mit der Gleichung

$$y = \int_a^b G(x, \xi)\, \tilde{f}\big(\xi, y(\xi)\big)\, d\xi = \tilde{T} y. \tag{4.14}$$

Die Ungenauigkeit von (4.13) sei in einem Bereich $\mathfrak{B}$ durch

$$\varrho\big(\tilde{f}(x, y(x)), f(x, y(x))\big) \leqq \mu \tag{4.15}$$

beschränkt. Dann ist

$$T y = \int_a^b G(x, \xi)\, f\big(\xi, y(\xi)\big)\, d\xi,$$

wobei $f(x, y)$ und also auch $T y$ nicht (genau) bekannt sind.

Wählen wir der Einfachheit halber als Abstand den Zahlenabstand

$$\varrho(u, v) = \|u(x) - v(x)\| = \max_{x \in \mathfrak{B}} |u(x) - v(x)|$$

und genügen die Funktionen $\tilde{f}$ und f in $\mathfrak{B}$ einer Lipschitzbedingung

$$|\tilde{f}(x, u) - \tilde{f}(x, v)| \leqq L |u - v|$$

* Vgl. L. Collatz [6], S. 180ff. und J. Schröder [10] §§ 3 und 4.

mit der gleichen Lipschitzkonstanten L, so können wir

$$P = L \max_{x \in \mathfrak{B}} \int_a^b |G(x, \xi)| \, d\xi \tag{4.16}$$

setzen und die Größe $\| Tu_0 - \tilde{T} u_0 \|$ durch

$$\| Tu_0 - \tilde{T} u_0 \| \leqq \mu \cdot \max_{x \in \mathfrak{B}} \int_a^b |G(x, \xi)| \, d\xi \tag{4.17}$$

abschätzen.

4. Beispiel: Als spezielles Beispiel betrachten wir die nichtlineare Schwingungsaufgabe

$$y'' + 4y + y^3 + \gamma y^5 = \sin x, \qquad |\gamma| \leqq 0,1, \tag{4.18}$$

wobei also γ nicht genau gegeben sei. Gesucht sei eine periodische Lösung $y(x)$ mit der Periode 2π der „erzwingenden Kraft" $\sin x$.

Hier existiert offenbar für die Randwertaufgabe

$$y'' + 4y = r(x), \qquad y(0) = y(2\pi), \qquad y'(0) = y'(2\pi),$$

keine Greensche Funktion*. Wir betrachten daher statt dessen das Randwertproblem (4.18), $y(0) = 0$, $y'(\pi/2) = 0$. Diese Randbedingungen sind hinreichend** für die Periodizität der Lösung mit der Periode 2π. Als Greensche Funktion für die Randwertaufgabe

$$y'' + 4y = r(x), \qquad y(0) = y'\left(\frac{\pi}{2}\right) = 0$$

findet man

$$G(x, \xi) = \begin{cases} -\frac{1}{2} \cos 2x \sin 2\xi & \text{für} \quad \xi \leqq x \\ -\frac{1}{2} \sin 2x \cos 2\xi & \text{für} \quad \xi \geqq x. \end{cases} \tag{4.19}$$

Wir iterieren nach der Vorschrift der Näherungsgleichung

$$y''_{n+1} + 4y_{n+1} = -y_n^3 + \sin x, \qquad y(0) = y'\left(\frac{\pi}{2}\right) = 0, \tag{4.20}$$

und beginnen mit der Lösung $y_0 = \frac{1}{3} \sin x$ der linearen Gleichung. Man erhält auf diese Weise die Näherungslösungen in der Gestalt

$$y_n = \sum_{\nu = 1, 3, 5, \ldots}^{k_n} a_\nu \sin \nu x, \tag{4.21}$$

wobei also nur Koeffizienten mit ungeradem Index auftreten. Dies gilt auch, wenn noch das Glied $-\gamma y^5$ hinzugenommen würde***.

---

* Für $r(x) = \sin 2x$ ist das Problem nicht lösbar. Vgl. auch L. COLLATZ [6] S. 25.

** H. EHRMANN [18] S. 133. Dort ist auch gezeigt, daß die Differentialgleichung (4.18) im Falle $\gamma \geqq 0$ unendlich viele periodische Lösungen mit der Periode 2π besitzt. Im Falle $\gamma < 0$ folgt die Existenz mindestens einer Lösung der Randwertaufgabe (4.18), $y(0) = y'(\pi/2) = 0$ und damit mindestens einer periodischen Lösung aus einem Existenzsatz von H. EPHESER [16] S. 451f.

*** Für y_0 ist dies richtig, und setzt man (4.21) in die rechte Seite $f(y_n(x), x) = g(x)$ von (4.20) ein, so gilt wegen der Symmetrie von $f(y, x)$ $g(-x) = -g(x)$ und $g(x + \pi) = -g(x)$. Die Fourierentwicklung von $g(x)$ enthält daher nur Glieder mit $\sin(2r + 1)x$, $r = 0, 1, 2, \ldots$. Siehe R. ZURMÜHL [17] S. 313f. Diese bleiben bis auf die Koeffizienten a_ν bei der Integration (4.20) erhalten.

Man erhält mit $y_0 = \frac{1}{3}\sin x$

$$y_1 = \frac{35}{108}\sin x - \frac{1}{540}\sin 3x = 0{,}324074074\sin x - 0{,}001851852\sin 3x,$$

$$y_2 = \frac{4091233}{12597120}\sin x - \frac{692891}{393660000}\sin 3x + \frac{11}{1574640}\sin 5x -$$

$$- \frac{7}{377913600}\sin 7x + \frac{1}{48498912000}\sin 9x$$

$$= 0{,}324775266\sin x - 0{,}001760125\sin 3x + 0{,}6986\cdot 10^{-5}\sin 5x -$$

$$- 1{,}85\cdot 10^{-8}\sin 7x + 0{,}2\cdot 10^{-10}\sin 9x.$$

Zur Durchführung der Fehlerabschätzung geben wir den Bereich

$$\mathfrak{B}: \quad 0 \leqq x \leqq \frac{\pi}{2}, \quad \|u - y_2\| \leqq 1\cdot 10^{-3}$$

vor. Man findet $\max_{\mathfrak{B}} |u| < 0{,}3276 = u_m$ und für die Lipschitzkonstante L das Maximum der Ableitungen von $\tilde{f}$ bzw. f nach u in $\mathfrak{B}$:

$$3u_m^2 + |\gamma|\cdot 5u_m^4 < 0{,}328 = L.$$

Weiter ergibt sich für die Greensche Funktion (4.19) nach leichter Rechnung die Abschätzung

$$\int_0^{\pi/2} |G(x,\xi)|\, d\xi \leqq \tfrac{1}{4}\int_0^{\pi/2} [|\sin 2(x+\xi)| + |\sin 2(x-\xi)|]\, d\xi = \tfrac{1}{2}.$$

Damit wird $P = 0{,}164$ und mit $\|y_1 - y_2\| < 0{,}0008$ sowie

$$\|f(x,u) - \tilde{f}(x,u)\| = \|\gamma u^5\| \leqq 0{,}1\cdot u_m^5 < 0{,}0003774$$

ergibt (3.25) für die periodische Lösung $y(x)$ der Schwingungsgleichung (4.18) den Fehler

$$\|y - y_2\| = \max_{0\leq x\leq \pi/2} |y(x) - y_2(x)| \leqq 0{,}000157 + 0{,}000452 = 0{,}000609.$$

Es liegt y in $\mathfrak{B}$. Daher ist die Abschätzung gültig, und in $\mathfrak{B}$ existiert genau diese eine periodische Lösung $y(x)$ bei beliebigem $|\gamma| \leqq 0{,}1$.

Literatur

[1] Kurepa, G.: Tableaux ramifiés d'ensembles. Espaces pseudo-distanciés. C. R. Acad. Sci. Paris **198**, 1563–1565 (1934).

[2] Kantorovitch, L.: The method of successive approximation for functional equations. Acta Math. **71**, 63–97 (1939).

[3] Weissinger, J.: Über das Iterationsverfahren. Z. angew. Math. Mech. **31**, 245–246 (1951).

[4] Weissinger, J.: Zur Theorie und Anwendung des Iterationsverfahrens. Math. Nachr. **8**, 193–212 (1952).

[5] Collatz, L.: Fehlerabschätzungen zum Iterationsverfahren bei linearen und nichtlinearen Randwertaufgaben. Z. angew. Math. Mech. **33**, 116–127 (1953).

[6] Collatz, L.: Numerische Behandlung von Differentialgleichungen, 2. Aufl. Berlin 1955.

[7] Collatz, L.: Einige funktionalanalytische Methoden bei der numerischen Behandlung von Differentialgleichungen. Vortrag Jahrestagung der Gesellschaft für Angew. Math. Mech., April 1958. Z. angew. Math. Mech. **38**, 264—267 (1958).
[8] Collatz, L.: Näherungsverfahren höherer Ordnung für Gleichungen in Banachräumen. Arch. Rational Mech. Anal. **2**, 66—75 (1958).
[9] Schröder, J.: Das Iterationsverfahren bei allgemeinerem Abstandsbegriff. Math. Z. **66**, 111—116 (1956).
[10] Schröder, J.: Neue Fehlerabschätzungen für verschiedene Iterationsverfahren. Z. angew. Math. Mech. **36**, 168—181 (1956).
[11] Schröder, J.: Nichtlineare Majoranten beim Verfahren der schrittweisen Näherung. Arch. Math. **7**, 471—484 (1956).
[12] Schröder, J.: Über das Newtonsche Verfahren. Arch. Rational Mech. Anal. **1**, 154—180 (1957).
[13] Carathéodory, C.: Variationsrechnung und partielle Differentialgleichungen erster Ordnung, Bd. I, 2. Aufl., herausgeg. von E. Hölder. Leipzig 1956.
[14] Vaughan, D.: On the form of satellites revolving at small distances from their primaries. Phil. Mag. **20** (1860).
[15] Unger, H.: Nichtlineare Behandlung von Eigenwertaufgaben. Z. angew. Math. Mech. **30**, 281—282 (1950).
[16] Epheser, H.: Über die Existenz der Lösungen von Randwertaufgaben mit gewöhnlichen, nichtlinearen Differentialgleichungen zweiter Ordnung. Math. Z. **61**, 435—454 (1955).
[17] Zurmühl, R.: Praktische Mathematik, 2. Aufl. Berlin 1957.
[18] Ehrmann, H.: Nachweis periodischer Lösungen bei gewissen nichtlinearen Schwingungsdifferentialgleichungen. Arch. Rational Mech. Anal. **1**, 124—138 (1957).

Mathematisches Institut A
der Technischen Hochschule
Stuttgart

(Eingegangen am 13. Juni 1959)

*Konstruktion und Durchführung von Iterationsverfahren höherer Ordnung**

HANS EHRMANN

Vorgelegt von L. COLLATZ

§ 1. Einleitung

Das Problem der Aufstellung von Iterationsverfahren für Gleichungen mit einer Zahl als Unbekannte, die mit einer vorgegebenen Ordnung konvergieren, wurde von zahlreichen Mathematikern immer wieder erneut in Angriff genommen[1]. Seine Lösung geht jedoch bereits auf LEONHARD EULER zurück[2]. Eine geschlossene vollständige Darstellung des unter gewissen Differenzierbarkeitsvoraussetzungen allgemeinsten Iterationsverfahrens von gegebener ganzzahliger Ordnung bringt E. SCHRÖDER 1870 [*1*]. Wenn hier trotzdem noch einmal dieser Problemkreis angeschnitten wird, obwohl das Problem der Aufstellung von Iterationsverfahren bestimmter Ordnung theoretisch vollständig gelöst ist, so geschieht dies, weil die meisten dieser Verfahren für die praktische Durchführung zu umständlich und unübersichtlich sind, so daß man sich bisher in der Praxis im allgemeinen auf die mehrfache Anwendung einfacher Verfahren mit kleinerer Konvergenzordnung beschränkte. Das Fehlen einer für die Rechnung handlichen Form hat offenbar auch dazu beigetragen, daß die Verfahren höherer Ordnung immer wieder in Vergessenheit geraten sind. Außerdem besteht meines Wissens trotz verschiedener Auszählungen der Rechenschritte[3] noch keine einfache Entscheidungsmöglichkeit, welche Verfahren in speziellen Fällen am günstigsten sind.

So wird beispielsweise in der praktischen Mathematik zur Bestimmung der Wurzeln algebraischer Gleichungen unter den hier betrachteten Iterationsverfahren in den meisten Fällen das Newtonsche Verfahren, also ein Verfahren 2. Ordnung, bevorzugt. In der Tat erweist es sich für algebraische Gleichungen bis zum 5. Grade den Verfahren höherer Ordnung überlegen. Dagegen ist für algebraische Gleichungen vom 6. Grade an das Verfahren 3. Ordnung in der in § 3 beschriebenen Form in Bezug auf die Rechenarbeit im allgemeinen allen anderen Verfahren der Ordnung $k>1$ vorzuziehen, was im Folgenden noch genauer erklärt wird.

* Diese Ausführungen sind ebenfalls wie eine frühere Veröffentlichung „Iterationsverfahren mit veränderlichen Operatoren" meiner Habilitationsschrift entnommen. Auch hier möchte ich den Professoren L. COLLATZ, H. KÖNIG und G. SCHULZ für ihre freundliche Unterstützung herzlich danken.

[1] Vgl. u. a. das Literaturverzeichnis am Schluß dieser Arbeit.

[2] Siehe E. BODEWIG [*2*].

[3] Siehe z. B. R. LUDWIG [*3*].

§ 2 ist ebenfalls nur als eine Einführung gedacht und bringt neben einigen Voraussetzungen und einfachen Folgerungen einige schon bekannte Ergebnisse (vgl. [1]), die hier nur etwas präzisiert werden. Die Ausführungen verfolgen im wesentlichen den Zweck, darauf hinzuweisen, daß die immer wieder unabhängig voneinander aufgestellten speziellen Iterationsverfahren höherer Ordnung sich unmittelbar aus allgemeineren Untersuchungen ergeben, die bereits E. SCHRÖDER durchgeführt hat, und daß die Aufstellung neuer solcher Verfahren keinerlei Schwierigkeiten macht.

In § 3 wird trotzdem eine neue Methode zur Aufstellung von speziellen Verfahren jeder ganzzahligen Ordnung $k > 1$ angegeben, weil es gelingt, diese speziellen Verfahren in eine Form zu bringen, die einerseits für jede Ordnung $k \geqq 3$ die praktische Durchführung gegenüber den bisherigen Methoden wesentlich vereinfacht[1], andererseits eine unmittelbare Abschätzung der Rechenarbeit gestattet. Überhaupt scheint bei diesen Verfahren die Rechenarbeit (in bezug auf Multiplikationen und Divisionen) die untere Grenze zu bilden von allen entsprechenden Verfahren derselben Ordnung, wie mehrere Vergleiche zeigten, ich jedoch leider nicht allgemein beweisen konnte.

In § 4 werden einige Sätze über die Rechenarbeit und die Wahl der Ordnung für die in § 3 aufgestellten Verfahren bewiesen, die eine Entscheidung gestatten, welche Ordnung bei speziellen Gleichungen die zweckmäßige ist.

§ 5 bringt Fehlerabschätzungen für die betreffenden Iterationsverfahren und § 6 schließlich einige numerische Beispiele.

Es sei noch erwähnt, daß sich die aufgestellten Verfahren höherer Ordnung besonders gut für eine Übertragung auf allgemeinere Gleichungen z. B. im Banach-Raum eignen, da jeweils nur *eine* Division $[(F'(x_n))^{-1}]$ erforderlich ist, die dann der allgemeinen Lösung einer linearen Gleichung $[T_n' u = v$ z. B. mit der Fréchetschen Ableitung $T_n']$ entspricht. Dies soll an anderer Stelle gezeigt werden.

§ 2. Voraussetzungen und einfache Folgerungen

Zur Bestimmung einer Lösung[2] ξ der Gleichung

$$F(x) = 0 \tag{2.1}$$

betrachten wir Iterationsverfahren der Gestalt[3]

$$x_{n+1} = f(x_n) = G\left(x_n, F(x_n), F'(x_n), \ldots, F^{(s)}(x_n)\right), \quad s \geqq 1, \quad n = 0, 1, 2, \ldots . \tag{2.2}$$

Ein solches Verfahren konvergiert für alle Anfangswerte x_0 einer Umgebung von ξ mindestens mit der Ordnung[4] $k > 1$ gegen ξ, wenn

$$f(x) - \xi = O(|x - \xi|^k), \qquad k > 1 \tag{2.3}$$

[1] Bei Rechnung mit elektronischen Maschinen gestatten sie u. a. wegen der einheitlichen Gestalt eine einfache geschlossene Programmierung für jedes $k \geqq 2$.

[2] Die Existenz von ξ setzen wir zunächst voraus.

[3] Wir schließen damit alle Verfahren aus, bei denen im Ausdruck für $f(x)$ iterierte Funktionen wie z. B. $F(F(x))$ u. a. vorkommen.

[4] Der Fall $k = 1$ macht weitere Fallunterscheidungen nötig. Wir beschränken uns hier auf überlineare Konvergenz.

gilt. Diese Bedingung ist erfüllt, wenn $f(x)$ in einer Umgebung von ξ Ableitungen bis zur k-ten Ordnung besitzt und die Gleichungen

$$f(\xi)=\xi, \quad f'(\xi)=f''(\xi)=\cdots=f^{(k-1)}(\xi)=0, \quad k>1, \tag{2.4}$$

erfüllt sind. Ist außerdem $f^{(k)}(\xi)\neq 0$, so hat das Verfahren genau die Ordnung[1] k.

Die folgenden einfachen Sätze sind im wesentlichen bekannt:

Satz 1. *Ergibt $f(x)$ in (2.2) ein Iterationsverfahren von der Ordnung $k>1$, so hat das Iterationsverfahren*

$$x_{n+1}=f_r(x_n), \quad n=0,1,2,\ldots,$$

wobei $f_r(x)$ die „r-fach iterierte" Funktion[2] ist, die Ordnung k^r.

Betrachtet man also $r>1$ Schritte des Verfahrens (2.2) als einen einzigen, so erhöht sich damit die Ordnung $k>1$ auf die r-te Potenz.

Satz 2. *Für die Funktionen $f(x)$ und $g(x)$ gelte*

$$f(x)-\xi=O(|x-\xi|^{k_1}) \quad \text{und} \quad g(x)-\xi=O(|x-\xi|^{k_2})$$

mit $k_1>1$ und $k_2\geqq 1$.

Dann konvergieren die Iterationsverfahren

$$x_{n+1}=f(g(x_n)) \quad \text{und} \quad x_{n+1}=g(f(x_n)), \quad n=0,1,2,\ldots,$$

für alle x_0 einer Umgebung von ξ mindestens mit der Ordnung $k_1\cdot k_2$.

Satz 3. *Ergibt $f(x)=s_k(x)$ in (2.2) ein Verfahren $(k>1)$-ter Ordnung gegen ξ, so erhält man mit $f^*(x)=s_k(x)+g(x)$ das allgemeinste[3] Verfahren k-ter Ordnung, wenn $g(x)$ eine (in einer Umgebung von ξ eindeutig definierte) Funktion ist, die lediglich der Bedingung*

$$g(x)=O(|x-\xi|^k) \tag{2.5}$$

unterworfen ist.

Satz 4. *Ergibt $f(x)=s_{k_1}(x)$ ein Iterationsverfahren, das mit der Ordnung $k_1>k>1$ gegen ξ konvergiert, so ist*

$$\text{im Falle } F'(\xi)\neq 0 \quad x_{n+1}=s_{k_1}(x_n)+|F(x_n)|^k$$

$$\text{im Falle } F'(\xi)=0 \quad x_{n+1}=s_{k_1}(x_n)+\left|\frac{F(x_n)}{F'(x_n)}\right|^k$$

ein Verfahren von genau k-ter Ordnung[4].

Beweis. Sowohl $F(x)$ im Falle $F'(\xi)\neq 0$ als auch $\dfrac{F(x)}{F'(x)}$ haben an der Stelle ξ eine einfache Nullstelle. Daher ist

$$|F(x)|^k \quad \text{bzw.} \quad \left|\frac{F(x)}{F'(x)}\right|^k \quad \text{gleich} \quad O(|x-\xi|^k),$$

[1] Diese Ordnungsdefinition geht auf E. Schröder zurück.

[2] $f_1(x)=f(x)$, $f_2(x)=f(f(x))$, …, $f_r(x)=f(f_{r-1}(x))$: „r-fach iterierte Funktion".

[3] Im Sinne von (2.3).

[4] Es ist also keine wesentliche Einschränkung, wenn wir uns im Folgenden auf die Konstruktion von Iterationsverfahren mit ganzzahliger Ordnung $k>1$ beschränken.

aber nicht $o(|x-\xi|^k)$. Zusammen mit

$$s_{k_1}(x) - \xi = O(|x-\xi|^{k_1}) = o(|x-\xi|^k)$$

folgt hieraus die Behauptung.

Haben wir bereits ein Verfahren k-ter Ordnung, $f(x) = s_k(x)$, so erhält man nach Satz 3 weitere solche Verfahren, wenn man zu $s_k(x)$ Funktionen $g(x)$ addiert, die (2.5) genügen, z.B. $g(x) = (x-\xi)^k$. Da aber ξ im allgemeinen nicht bekannt ist, müssen wir $g(x)$ durch die Funktion $F(x)$ und ihre Ableitungen ausdrücken, z.B. $g(x) = [F(x)]^k$ oder allgemeiner $g(x) = [F(x)]^k \varphi(x)$ mit einer Funktion $\varphi(x)$, die für $x \to \xi$ beschränkt bleibt. Unter gewissen einfachen Voraussetzungen über die Funktion $F(x)$ erhält man auf diese Weise auch *alle* Verfahren, die (mindestens) mit der Ordnung k gegen ξ konvergieren. Es gilt der

Satz 5. *Die Funktion $F(x)$ besitze in ξ eine Nullstelle, sei in einer Umgebung von ξ differenzierbar, und es sei $F'(\xi) \neq 0$.*

Ergibt dann (2.2) mit $f(x) = s_k(x)$, $k > 1$, ein Iterationsverfahren k-ter Ordnung, das gegen ξ konvergiert, so erhält man mit

$$f^*(x) = s_k(x) + [F(x)]^k \varphi(x) \tag{2.6}$$

mit einer willkürlichen Funktion $\varphi(x)$, die noch von $F(x)$ und ihren Ableitungen abhängen kann und die für $x \to \xi$ beschränkt bleibt, das allgemeinste[1] *Verfahren k-ter Ordnung, bei dem im Ausdruck für $f(x)$ die unbekannte Größe ξ nicht auftritt.*

Beispiel. Ersetzt man z.B. beim Newtonschen Verfahren $f(x) = x - \frac{F}{F'}$, das bekanntlich die Ordnung 2 nur bei einfacher Nullstelle ξ von $F(x)$ $[F'(\xi) \neq 0]$ hat, die Funktion F durch die Funktion $F_1 = \frac{F}{F'}$, so erhält man das Verfahren $f_1(x) = x - \frac{FF'}{F'^2 - FF''}$, das auch bei mehrfacher Nullstelle von $F(x)$ die Ordnung 2 hat. Aus Satz 5 folgt, daß sich die Funktionen $f(x)$ und $f_1(x)$ nur durch einen Summanden $F^2 \cdot \varphi(x)$ unterscheiden, wobei $\varphi(x)$ im Falle einer einfachen Nullstelle ξ von F für $x \to \xi$ beschränkt bleibt, im Falle einer mehrfachen Nullstelle aber nicht beschränkt bleiben kann[2]. Man erhält in der Tat eine solche Funktion: $\varphi(x) = \frac{F''}{F'(F'^2 - FF'')}$.

Beweis von Satz 5. a) Es ist nach dem Mittelwertsatz

$$F(x) = (x-\xi) F'(\tilde{x}), \qquad \tilde{x} \text{ Zwischenwert}, \tag{2.7}$$

wobei $F'(\tilde{x})$ nach Voraussetzung in einer Umgebung von ξ beschränkt und $\neq 0$ ist. Hieraus folgt

$$f^*(x) - \xi = s_k(x) - \xi + [F(x)]^k \varphi(x) = O(|x-\xi|^k) + O(|x-\xi|^k) = O(|x-\xi|^k).$$

Das Verfahren $f^*(x)$ konvergiert also mit der Ordnung $k > 1$.

b) Andererseits muß nach Satz 3 die Funktion $f^*(x)$ die Gestalt

$$f^*(x) = s_k(x) + g(x) \quad \text{mit} \quad g(x) = O(|x-\xi|^k)$$

[1] Es werden wieder nur solche Iterationsverfahren betrachtet, für die (2.3) gilt.
[2] Das F_1 nur eine einfache Nullstelle hat, ist Satz 5 hier anwendbar.

haben. Es muß also der Quotient

$$\frac{|g(x)|}{|x-\xi|^k} \quad \text{für} \quad x \to \xi$$

beschränkt bleiben. Wir können daher $g(x)$ in der Form

$$g(x) = (x-\xi)^k \psi(x) \tag{2.8}$$

mit einer für $x \to \xi$ beschränkten, aber sonst willkürlichen Funktion $\psi(x)$ schreiben. Wegen $F'(\tilde{x}) \neq 0$ können wir

$$\varphi(x) = \frac{\psi(x)}{[F'(\tilde{x})]^k} \tag{2.9}$$

setzen, ohne die willkürliche Wahl von $\varphi(x)$ einerseits bzw. von $\psi(x)$ andererseits einzuschränken[1]. Mit $\psi(x)$ bleibt auch $\varphi(x)$ für $x \to \xi$ beschränkt und umgekehrt. Aus (2.7), (2.8) und (2.9) folgt für $g(x)$ die Form

$$g(x) = [F(x)]^k \varphi(x), \qquad \text{w.z.b.w.}$$

Zusatz. Ist $F'(\xi) = 0$, hat also $F(x)$ an der Stelle ξ eine mehrfache z. B. p-fache Nullstelle $\left(F'(\xi) = F''(\xi) = \cdots = F^{(p-1)}(\xi) = 0,\ F^{(p)}(\xi) \neq 0\right)$, so bleibt der Satz gültig und der Beweis unverändert, wenn man überall $F(x)$ durch $\frac{F(x)}{F'(x)}$ oder durch $F^{(p-1)}(x)$ ersetzt.

Satz 4 und Satz 5 zusammen zeigen, daß man *alle* Verfahren, die mit einer vorgegebenen Ordnung $k > 1$ gegen eine Lösung ξ von (2.1) konvergieren und für die (2.3) gilt, erfaßt, wenn man *ein* solches Verfahren für jedes ganzzahlige $k > 1$ angeben kann. Dieses Problem ist, wie schon bemerkt, bereits vollständig gelöst. In geschlossener Form geben E. Schröder [*1*] und später E. Bodewig [*2*] für ein Verfahren k-ter Ordnung ($k \geqq 2$) im Falle $F'(\xi) \neq 0$ die Formel an

$$f(x) = x + \sum_{\nu=1}^{k-1} (-1)^\nu \frac{F^\nu}{\nu!} \left(\frac{1}{F'} \frac{d}{dx}\right)^{\nu-1} \frac{1}{F'} \qquad \text{E. Schröder 1869,} \tag{2.10}$$

wobei $\left(\frac{1}{F'} \frac{d}{dx}\right)^r$ bedeutet, daß der Operator $\frac{1}{F'} \frac{d}{dx}$ r-mal anzuwenden ist, also z. B. $\left(\frac{1}{F'} \frac{d}{dx}\right)^3 g(x) = \frac{1}{F'(x)} \frac{d}{dx} \left[\frac{1}{F'(x)} \frac{d}{dx} \left(\frac{1}{F'(x)} g'(x)\right)\right]$.

Bei einer mehrfachen Nullstelle ξ $\left(F'(\xi) = 0\right)$ kann man auch hier in (2.10) überall $F(x)$ durch $H(x) = \frac{F(x)}{F'(x)}$ ersetzen, um die Ordnung k zu sichern.

Da die Formeln mit $F(x)$ statt $H(x)$ im allgemeinen jedoch wesentlich einfacher sind, setzen wir für das Folgende $F'(\xi) \neq 0$ voraus.

Für die praktische Durchführung der Verfahren werden die Ausdrücke (2.10) aber sehr bald zu unübersichtlich[2]. Es ist daher für die Praxis von großer

[1] Man beachte, daß $F'(\tilde{x})$ bei festem ξ eine Funktion von x allein ist.

[2] Zum Beispiel erhält man für $k = 4$ aus (2.10) den Ausdruck

$$f(x) = x - \frac{F}{F'} - \frac{F^2 F''}{2F'^3} - \frac{F^3}{6F'^4}\left(3\frac{F''^2}{F'} - F'''\right).$$

Eine zweimalige Durchführung des Newtonschen Verfahrens erscheint hier im allgemeinen vorteilhafter.

Bedeutung, Iterationsverfahren anzugeben, die mit höherer Ordnung ($k>2$) konvergieren und deren Durchführung praktisch noch sinnvoll ist. Ferner wird eine einfache Entscheidungsmöglichkeit verlangt, welche Ordnung k in speziellen Fällen die zweckmäßige ist.

Mit diesen Problemen werden wir uns in den nächsten beiden Paragraphen beschäftigen.

§ 3. Aufstellung und Durchführung der Verfahren

Es gilt der

Hilfssatz 1. *Die nach beiden Argumenten genügend oft differenzierbaren Funktionen* $g_k(x, \xi)$ *und* $g_{k+1}(x, \xi)$ *ergeben gegen* ξ *konvergente Iterationsverfahren*

$$x_{n+1} = g_k(x_n, \xi) \qquad \textit{von } k\textit{-ter Ordnung, } k>1, \textit{ ganz,}$$

$$\textit{bzw.} \quad x_{n+1} = g_{k+1}(x_n, \xi) \qquad \textit{von } (k+1)\textit{-ter Ordnung.}$$

Ferner gelte

$$\left.\frac{\partial g_{k+1}(x, \xi)}{\partial \xi}\right|_{x=\xi} = 0. \tag{3.1}$$

Dann konvergiert das Iterationsverfahren

$$x_{n+1} = g_{k+1}\big(x_n, g_k(x_n, \xi)\big), \qquad n = 0, 1, 2, \ldots,$$

(mindestens) mit $(k+1)$*-ter Ordnung gegen* ξ.

Der einfache Beweis ergibt sich unmittelbar durch Nachprüfen der Gln. (2.4) mit $f(x) = g_{k+1}\big(x, g_k(x, \xi)\big)$.

Durch Entwicklung von $F(\xi)$ an einer Nachbarstelle x von ξ in eine Taylor-Reihe ergibt sich

$$0 = F(\xi) = F(x) + (\xi - x)\,F'(x) + \frac{(\xi-x)^2}{2!}F''(x) + \cdots. \tag{3.2}$$

Wir setzen zunächst die Konvergenz dieser Reihe in einer Umgebung U von ξ voraus. Ferner sei in U $F'(x) \neq 0$. Dann können wir (3.2) in der Form schreiben:

$$\xi = x - \frac{F}{F'} - \frac{1}{F'}\left\{\frac{(\xi-x)^2}{2!}F'' + \frac{(\xi-x)^3}{3!}F''' + \cdots\right\}, \qquad (F' \neq 0). \tag{3.3}$$

Ersetzen wir *links* ξ durch x_{n+1} und auf der rechten Seite x durch x_n, so erhalten wir mit

$$x_{n+1} = x_n - \frac{F_n}{F_n'} - \frac{1}{F_n'}\left\{\frac{(\xi-x_n)^2}{2!}F_n'' + \frac{(\xi-x_n)^3}{3!}F_n''' + \cdots\right\} \tag{3.4}$$

ein Verfahren, das mit dem 1. Schritt die Lösung ξ von $F(x)=0$ liefert, aber natürlich praktisch wegen des Auftretens der Größe ξ und der unendlich vielen Glieder keinen Sinn hat. Brechen wir die Reihe (3.4) nach dem 2. Gliede $\left(\frac{-F_n}{F_n'}\right)$ ab, so haben wir wieder das Newtonsche Verfahren. Brechen wir sie nach dem Gliede mit $(\xi - x_n)^k$, $k \geqq 2$, ab, so erhalten wir ein Verfahren (mindestens) $(k+1)$-ter Ordnung. In diesem Falle genügt es, wenn $F(x)$ $(k+1)$-mal nach x differenzierbar ist:

Satz 6. *Es sei $F(x)$ in einer Umgebung U einer Nullstelle ξ $(k+1)$-mal differenzierbar, $k \geqq 1$. Ferner sei dort $F'(x) \neq 0$.*

Dann ist $x_{n+1} = g_{k+1}(x_n, \xi)$ mit

$$(3.5)\quad g_{k+1}(x,\xi) = x - \frac{F(x)}{F'(x)} - \frac{1}{F'(x)}\left\{\frac{(\xi-x)^2}{2!}F''(x) + \cdots + \frac{(\xi-x)^k}{k!}F^{(k)}(x)\right\},\quad k \geqq 2$$

ein Iterationsverfahren $(k+1)$-ter Ordnung.

Beweis. Bricht man die Reihe in (3.3) nach dem $(k+1)$-ten Gliede unter Berücksichtigung des Restgliedes ab und subtrahiert dann (3.3) von (3.5), so ergibt sich

$$g_{k+1}(x) - \xi = \frac{1}{F'(x)}\,\frac{(\xi-x)^{k+1}}{(k+1)!}F^{(k+1)}(\tilde{x}),$$

wobei $\tilde{x}$ ein Zwischenwert ist, der in einem Bereich liegt, der x und ξ enthält. Daraus folgt $g_{k+1}(x) - \xi = O(|x-\xi|^{k+1})$, w.z.z.w.

Nun ist (3.5) wegen des Auftretens von ξ in dieser Form praktisch unbrauchbar. Haben wir aber bereits ein Verfahren k-ter Ordnung, $x_{n+1} = g_k(x_n)$, in dem der Wert ξ nicht vorkommt, so erhalten wir nach dem obigen Hilfssatz ein Verfahren $(k+1)$-ter Ordnung, wenn wir in (3.5) ξ durch $g_k(x)$ ersetzen, und in diesem Verfahren tritt dann ebenfalls die Größe ξ nicht mehr auf. Auf diese Weise ist es möglich, rekursiv Verfahren beliebig hoher Ordnung zu berechnen nach der Vorschrift[1]:

$$(3.6)\quad \begin{aligned} f_2(x) &= x - \frac{F(x)}{F'(x)} \\ f_{k+1}(x) &= x - \frac{F(x)}{F'(x)} - \frac{1}{F'(x)}\left\{\frac{(f_k(x)-x)^2}{2!}F''(x) + \cdots + \frac{(f_k(x)-x)^k}{k!}F^{(k)}(x)\right\}, \\ &\qquad k \geqq 2. \end{aligned}$$

Man erhält z. B. auf diese Weise

$$(3.7)\quad \begin{aligned} f_3(x) &= x - \frac{F}{F'} - \frac{F^2F''}{2F'^3}, \\ f_4(x) &= x - \frac{F}{F'} - \frac{1}{F'}\left\{\frac{F''}{2}\left(\frac{F}{F'} + \frac{F^2F''}{2F'^3}\right)^2 - \frac{F'''}{6}\left(\frac{F}{F'} + \frac{F^2F''}{2F'^3}\right)^3\right\} \quad \text{usw.}^2. \end{aligned}$$

Der Vorzug von (3.6) liegt jedoch nicht in der Einfachheit und Übersichtlichkeit der Methode zur Aufstellung von Iterationsverfahren k-ter Ordnung ($k > 1$, ganz), sondern darin, daß (3.6) die Möglichkeit bietet, die Durchführung eines solchen Verfahrens k-ter Ordnung für die praktische Rechnung sehr zu erleichtern[3]. Der Grundgedanke hierfür besteht darin, daß es für die Anwendung eines Verfahrens k-ter Ordnung nicht erforderlich ist, das Verfahren wie etwa

---

[1] Vgl. auch R. Zurmühl [5] S. 17.

[2] Multipliziert man die Klammern aus, so könnten jeweils noch die Glieder mit F^k und den höheren Potenzen von F vernachlässigt werden, ohne daß sich die Ordnung k ändert, wie aus Satz 5 folgt.

[3] Dies ist auch der einzige Grund, warum hier die zahlreichen Methoden zur Aufstellung von Iterationsverfahren k-ter Ordnung noch um eine weitere vermehrt werden.

in der Form (3.7) für $k=3$ oder 4 explizit anzugeben und die einzelnen Rechenoperationen an Hand dieser Formeln auszuführen, sondern man kann die obige Herleitung eines solchen Verfahrens in der Rechnung selbst vollziehen:

Satz 7. *Die Funktion $F(x)$ sei in der Umgebung einer Nullstelle ξ mindestens $(k+1)$-mal stetig differenzierbar mit $k \geqq 1$, und es sei $F'(\xi) \neq 0$.*

Dann ergibt folgende Vorschrift ein Iterationsverfahren, das mindestens mit der Ordnung $k+1$ gegen ξ konvergiert[1]*:*

Es sei x_n ein Näherungswert von ξ. Dann berechne man

1. *die Funktionswerte*

$$\frac{1}{l!}\frac{d^l F(x_n)}{dx^l} = \frac{F_n^{(l)}}{l!} \quad \text{für} \quad l = 0, 1, \ldots, k,$$

2. *die Verbesserungen*

$$v_{n,2} = \frac{-F_n}{F_n'}, \qquad v_{n,3} = \frac{-1}{F_n'}\left\{F_n + v_{n,2}^2 \frac{F_n''}{2!}\right\}, \ldots,$$

$$v_{n,k+1} = \frac{-1}{F_n'}\left\{F_n + v_{n,k}^2 \frac{F''}{2!} + v_{n,k}^3 \frac{F_n'''}{3!} + \cdots + v_{n,k}^k \frac{F_n^{(k)}}{k!}\right\}.$$

So wird der neue Näherungswert

$$x_{n+1} = x_n + v_{n,k+1}. \tag{3.8}$$

Beweis. Das Newtonsche Verfahren, $x_{n+1} = h(x_n) = x_n - \frac{F_n}{F_n'}$, ist ein Verfahren 2. Ordnung (wegen $F'(\xi) \neq 0$), denn man erhält durch Abbrechen der Reihe (3.2)

$$h(x) - \xi = \frac{-F''(\tilde{x})}{2F'(x)}(\xi - x)^2 = O(|x - \xi|^2), \tag{3.9}$$

wobei $\tilde{x}$ ein Zwischenwert ist, der in einem Bereich liegt, der x und ξ enthält.

Ist $x_{n+1} = x_n + v_{n,r}$ ein Verfahren r-ter Ordnung für $2 \leqq r \leqq k$, so folgt aus Satz 6 und Hilfssatz 1, daß

$$x_{n+1} = x_n - \frac{1}{F_n'}\left\{F_n + \frac{v_{n,r}^2}{2!} F_n'' + \cdots + \frac{v_{n,r}^r}{r!} F_n^{(r)}\right\}$$

ein Verfahren ist, das mindestens von $(r+1)$-ter Ordnung konvergiert. Setzt man r der Reihe nach $2, 3, \ldots, k$, so ergibt sich die Behauptung.

Bemerkung. Zur Durchführung des Verfahrens wird offenbar nur k-malige Differenzierbarkeit von $F(x)$ in ξ verlangt. In diesem Fall kann jedoch nicht die Ordnung $k+1$ gesichert werden[2].

Für die praktische Durchführung ist wesentlich, daß die Verbesserungen $v_{n,r+1}$, $r = 2, 3, \ldots, k$, ganze rationale Funktionen von $v_{n,r}$ sind, wobei die Koeffizienten $\frac{F^{(l)}}{l!}$ bezüglich r konstant bleiben. Nur der Grad ändert sich jeweils um Eins. Daher wird man hier zweckmäßig das Hornersche Schema anwenden.

[1] Alle vorkommenden Verfahren konvergieren jeweils in einer Umgebung von ξ, ohne daß es hier stets ausdrücklich erwähnt wird.

[2] Man kann leicht ein Gegenbeispiel angeben.

Eine passende Anordnung ist die der Fig. 1. Der Kürze halber ist dort für die Verbesserungen v_r anstatt $v_{n,r}$ geschrieben. *Als wesentlicher Punkt in dem Schema ist zu beachten, daß in der Folge der Koeffizienten* $c_l^{(l)} = \frac{F_n^{(l)}}{l!}$, $l = 0, 1, 2, \ldots, k$, *die Stelle für* $l = 1$, *also* F_n' *durch* 0 *zu ersetzen ist!*

Das Schema der Fig. 1 erweist sich besonders günstig, wenn $F(x)$ eine ganze rationale Funktion ist und man zur Berechnung der Funktionswerte $\frac{F^{(l)}}{l!}$ das gewöhnliche Hornersche Schema benutzt; denn in diesem Fall erscheinen diese Funktionswerte gerade an den richtigen Stellen.

§ 4. Über den Rechenaufwand und die Wahl der günstigsten Ordnung

Um eine Übersicht über den Rechenaufwand der einzelnen Verfahren zu bekommen, beschränken wir uns im Folgenden auf das Auszählen der Multiplikationen und Divisionen und betrachten diese beiden Rechenoperationen in bezug auf den Rechenaufwand als gleichwertig. Ferner untersuchen wir nur Iterationsverfahren (3.8) nach Satz 7 und Fig. 1.

Aus dem Schema folgt der

Satz 8. *Gelten die Voraussetzungen des Satzes 7 und sei M die Anzahl der Multiplikationen und Divisionen zusammen, so sind zur Durchführung eines Schrittes bei einem Verfahren k-ter Ordnung mit* $k \geqq 2$ *neben der Berechnung der k Funktionswerte*

$$F_n, \quad F_n', \quad \frac{F''}{2!}, \ldots, \quad \frac{F_n^{(k-1)}}{(k-1)!}, \quad \left[F_n^{(l)} = F^{(l)}(x_n)\right],$$

im allgemeinen Fall $\left(F^{(l)}(x_n) \neq 0,\ l = 1, 2, \ldots, k-1\right)$

$$M = \frac{k(k+1)}{2} - 2 \tag{4.1}$$

Multiplikationen bzw. Divisionen durchzuführen[1].

Der *Beweis* ergibt sich unmittelbar durch Auszählen aus dem Schema (Fig. 1).

Folgerungen. Im allgemeinen Fall steigt also der Rechenaufwand bezüglich der Multiplikationen bzw. Divisionen quadratisch mit der Ordnung k. Erhöht man stattdessen die Ordnung, indem man ein Verfahren $(k^* > 1)$-ter Ordnung, z. B. das Newtonsche, mehrfach anwendet, so steigt, wie sich aus Satz 1 ergibt, der Rechenaufwand bezüglich Multiplikationen und Divisionen nur wie $\log k$. Ähnlich verhält es sich auch mit der Anzahl s der zu berechnenden Funktionswerte $F_n, F_n', F_n'', \ldots$, die bei Erhöhung der Ordnung k im ersten Fall, d. h. bei einem Schritt mit einem Verfahren k-ter Ordnung mit k linear ansteigt, $s = k$, dagegen bei mehrfacher Anwendung eines Verfahrens $(k^* > 1)$-ter Ordnung nur logarithmisch mit k anwächst[2].

---

[1] Dabei ist die Berechnung von $b = \frac{-1}{F_n'}$ (s. Fig. 1) nicht mitgezählt, sondern es wurde jeweils $c_l b = \frac{-c_l}{F_n'}$, $l = 0, 1, \ldots, k-1$, als eine Division gezählt.

[2] Zum Beispiel ist bei Anwendung eines Verfahrens von der Ordnung k^* nach dem Satz 1

$$s = \frac{k^* \log k}{\log k^*}.$$

				$b = \frac{-1}{F_n'}$	$x_{n+1} = x_n + v_r$
			$c_0 = F_n$	$v_2 = c_0\, b$	Verfahren 2. Ordnung (NEWTON)
		0			
		$c_2'' v_2$	$c_2' v_2$		
	$c_2'' = \frac{F_n''}{2!}$	$c_2'' v_2 = c_2'$	$c_0 + c_2' v_2 = c_2$	$v_3 = c_2\, b$	Verfahren 3. Ordnung
	$c_3''' v_3$	$c_3'' v_3$	$c_3' v_3$		
$c_3''' = \frac{F_n'''}{3!}$	$c_2'' + c_3''' v_3 = c_3''$	$c_3'' v_3 = c_3'$	$c_0 + c_3' v_3 = c_3$	$v_4 = c_3\, b$	Verfahren 4. Ordnung
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
$c_k^{\mathrm{IV}} v_k$	$c_k''' v_k$	$c_k'' v_k$	$c_k' v_k$		
$c_k^{(k)} = \frac{F_n^{(k)}}{k!}$	$c_3''' + c_k^{\mathrm{IV}} v_k = c_k'''$	$c_2'' + c_k''' v_k = c_k''$	$c_k'' v_k = c_k'$	$c_0 + c_k' v_k = c_k$	$v_{k+1} = c_k\, b$	Verfahren $(k+1)$-ter Ordnung

Fig. 1. Rechenschema für ein Iterationsverfahren $(k+1)$-ter Ordnung $(k \geq 1)$

Man erkennt hieraus unmittelbar, daß es im allgemeinen Fall ($F_n^{(l)} \neq 0$ für $l = 1, 2, \ldots$) gewöhnlich sinnlos ist, die Ordnung eines Verfahrens beliebig zu steigern, da die dabei erzielte Genauigkeit durch einen unverhältnismäßig hohen Rechenaufwand erkauft werden muß, der bei mehrfacher Anwendung eines Verfahrens kleinerer Ordnung (> 1) geringer ist.

Unter gewissen einfachen Voraussetzungen läßt sich eine Aussage für die günstigste Ordnung machen:

Satz 9. *Es werde die Rechenarbeit für die Berechnung der Funktionswerte* $F(x)$, $F'(x)$, $\frac{F''(x)}{2!}, \ldots, \frac{F^{(k-1)}(x)}{(k-1)!}$ *an einer beliebigen Stelle* $x = x_n$ *in der Umgebung von* ξ *als gleich angesehen. Sie möge zur Berechnung eines dieser Funktionswerte der Rechenarbeit von a Multiplikationen bzw. Divisionen entsprechen.*

Dann ist für[1]

$$a < a^* = \frac{\log \frac{16}{3}}{\log \frac{9}{8}} \approx 14{,}213 \quad \text{das Newtonsche Verfahren } (k = 2)$$

und für

$$a > a^* \quad \text{das Verfahren 3. Ordnung}$$

unter allen Verfahren (3.8) *das günstigste.*

Bemerkung. Dieser Satz kann für die Fälle als Anhaltspunkt für die Wahl der Ordnung genommen werden, bei denen sich die Berechnung der Ableitungen von $F(x)$ nicht wesentlich gegenüber der von $F(x)$ selbst vereinfacht. Die Grenze a^* verschiebt sich in vielen Fällen noch etwas zugunsten des Verfahrens 3. Ordnung, also nach unten, wenn man berücksichtigt, daß die Durchführung des Verfahrens 3. Ordnung an Hand der Fig. 1 sehr einfach und übersichtlich möglich ist, was für die Berechnung von $F(x)$ und $F'(x)$ häufig nicht gilt.

Beweis von Satz 9. Die Anzahl der zu berechnenden Funktionswerte F, $\frac{F'}{1!}, \ldots, \frac{F^{(r-1)}}{(r-1)!}$ bei einem Verfahren r-ter Ordnung ($r > 1$) ist r. Bei n-maliger Anwendung müssen daher nr Funktionswerte berechnet werden, was nach Voraussetzung der Rechenarbeit von $s = anr$ Multiplikationen bzw. Divisionen entspricht. Dazu kommen nach (4.1) $m = n\left[\frac{r(r+1)}{2} - 2\right]$ solcher Rechenoperationen für die n-malige Durchführung des Verfahrens. Wir haben somit insgesamt die Arbeit

$$A = s + m = anr + \frac{n}{2}\left[r^2 + r - 4\right]$$

zu leisten für eine Gesamtordnung $k = r^n$, wie sie sich aus dem Satz 1 ergibt. Damit wird

$$A = \frac{\log k}{2 \log r}\left[r^2 + (2a+1)\, r - 4\right]. \tag{4.2}$$

Halten wir nun die Gesamtordnung $k \geqq 2$ fest[2], so wird die Arbeit A eine Funktion von der Ordnung r des angewandten Verfahrens allein: $A = A(r)$.

[1] Wir verstehen im Folgenden unter log stets den natürlichen Logarithmus.

[2] Wir lassen dabei auch solche ganzzahligen $k \geqq 2$ zu, die nicht in der Form r^n mit ganzzahligen r und n geschrieben werden können und geben in einer Ergänzung zu dem Beweis eine Begründung hierfür.

Es ist (wenn r zunächst stetig variabel angenommen wird)

$$\frac{dA}{dr} = \frac{\log k}{2} \cdot \frac{[2r^2+(2a+1)r](\log r - 1)+r^2+4}{r(\log r)^2}.$$

Hieraus folgt

$$\frac{dA}{dr} > 0 \qquad \text{für } r \geqq 3 \text{ (wegen } \log r > 1 \text{ für } r \geqq 3).$$

Daher ist stets

$$A(3) < A(r_1) \qquad \text{für } r_1 = 4, 5, \ldots .$$

Das bedeutet, daß im Durchschnitt[1] die Rechenarbeit zur Erzielung einer Gesamtordnung k für *alle* Verfahren (3.8) mit der Ordnung $r>3$ größer ist als bei Anwendung des Verfahrens der Ordnung 3. Es ist daher nur noch $r=2$ und $r=3$ zu vergleichen. Für diese Werte wird nach (4.2)

$$A(2) = \log k \, \frac{2a+1}{\log 2} \quad \text{und} \quad A(3) = \log k \, \frac{3a+4}{\log 3}.$$

Hieraus folgt

$$A(2) \lesseqgtr A(3) \quad \text{für} \quad a \lesseqgtr \frac{4\log 2 - \log 3}{2\log 3 - 3\log 2} = \frac{\log \frac{16}{3}}{\log \frac{9}{8}} - a^*.$$

Damit ist der Satz 9 bewiesen.

Ergänzung. Unter dem „günstigsten" Verfahren haben wir in Satz 9 das Verfahren r-ter Ordnung ($r \geqq 2$) verstanden, bei dem man zur Erzielung einer vorgegebenen Gesamtordnung k mit der kleinsten Rechenarbeit, d.h. mit der geringsten Anzahl von Multiplikationen und Divisionen auskommt. Theoretisch kann dabei k eine beliebige natürliche Zahl >1 sein. Da man bei der Rechnung aber nur stets eine ganzzahlige Anzahl n von Iterationsschritten durchführen kann und ebenfalls r eine ganze positive Zahl ($\geqq 2$) ist, so erhält man nur ganzzahlige Gesamtordnungen der Form $k=r^n$ bei n-maliger Durchführung eines Verfahrens r-ter Ordnung. Ist nun $r_1 \neq r_2$ und kommen z.B. in der Zerlegung von r_1 und r_2 in Primfaktoren verschiedene Primzahlen vor, so ist auch stets $k_1 = r_1^{n_1} \neq r_2^{n_2} = k_2$, wenn n_1 und n_2 natürliche Zahlen sind. Um zwei solche Verfahren mit den Ordnungen r_1 und r_2 zu vergleichen, haben wir im Beweis von Satz 9 auch nichtganzzahlige k zugelassen, die sich also praktisch nicht realisieren lassen. Für denjenigen, den dieser Weg über nichtrealisierbare Ordnungen k stört, sei zur Begründung hier folgender Satz angeführt:

Satz 10[2]. *Die Rechenarbeit (an Multiplikationen und Divisionen) zur Erzielung einer Gesamtordnung $k = r^n$ durch n-malige Durchführung eines Iterationsverfahrens von der Ordnung $r \geqq 2$ sei gegeben durch den Ausdruck*

$$A = Q(r) \log k, \tag{4.3}$$

wobei $Q(r)$ eine Funktion von r ist, die für (ganzzahlige) $r \geqq 2$ definiert ist und dort positive Werte annimmt.

[1] Siehe Fußnote 2, S. 75.

[2] Alle auftretenden kleinen lateinischen Buchstaben bezeichnen in diesem Satz natürliche Zahlen.

Es sei ferner für ein $r_1 \geqq 2$ *und alle für* $r_j \geqq 2$, $r_j \neq r_1$,

$$Q_1 = Q(r_1) < Q(r_j) = Q_j.$$

Dann folgt aus

$$k_1 = r_1^{n_1} \leqq r_j^{n_j} = k_j$$

stets

$$A_1 < A_j.$$

Dagegen existieren für jedes $j \neq 1$ *(natürliche) Zahlen* n_1 *und* n_j *derart, daß für*

$$k_1 = r_1^{n_1} > r_j^{n_j} = k_j$$

ebenfalls

$$A_1 < A_j$$

gilt.

Der Satz besagt, daß das Verfahren mit der Ordnung r_1 das günstigste ist in dem Sinne, daß 1. zur Erzielung derselben oder einer kleineren Gesamtordnung k eine kleinere Rechenarbeit erforderlich ist als bei allen anderen Verfahren $r_j \neq r_1$ und daß 2. auch mit geringerer Rechenarbeit eine größere Leistung, d.h. größere Gesamtordnung k möglich ist als bei Anwendung irgendeines Verfahrens $r_j \neq r_1$. Damit ist auch durch realisierbare Fälle der Begriff der günstigsten Ordnung r definiert. Da Gl. (4.2) die Gestalt (4.3) hat, ist damit unser Vorgehen im Beweis von Satz 9 hinreichend begründet.

Beweis von Satz 10. Aus $Q_1 < Q_j$ und $k_1 \leqq k_j$ folgt

$$A_1 = Q_1 \log k_1 < Q_j \log k_j = A_j.$$

Um den zweiten Teil der Behauptung zu beweisen, wählen wir eine rationale Zahl $\frac{n_1}{n_j}$ mit

$$\frac{Q_j \log r_j}{Q_2 \log r_1} > \frac{n_1}{n_j} > \frac{Q_1 + Q_j}{2 Q_1} \frac{\log r_j}{\log r_1},$$

was wegen $Q_j > Q_1 > 0$ und $r_1, r_j \geqq 2$ möglich ist. Hieraus folgt einerseits

$$\frac{n_1 \log r_1}{n_j \log r_j} = \frac{\log k_1}{\log k_j} > \frac{Q_1 + Q_j}{2 Q_1} > 1, \qquad \text{also } k_1 > k_j, \text{ andererseits}$$

$$\frac{A_j}{A_1} = \frac{Q_j n_j \log r_j}{Q_1 n_1 \log r_1} > 1, \qquad \text{w.z.b.w.}$$

Mit Satz 9 werden die Fälle nicht erfaßt, bei denen die Rechenarbeit zur Berechnung der Ableitungen $F^{(l)}(x)$ mit wachsendem l abnimmt. Dies ist insbesondere bei Polynomen der Fall. Für diese lassen sich leicht genaue Angaben über den Rechenaufwand machen. Es gilt der

Satz 11. *Bei der Lösung algebraischer Gleichungen n-ten Grades durch Iteration nach (3.8) ist in bezug auf die Rechenarbeit (Anzahl der Multiplikationen und Divisionen) bis zum Grade* $n = 5$ *das Newtonsche Verfahren und für einen Grad* $n > 5$ *das Verfahren 3. Ordnung unter allen Verfahren* $(r \geqq 2)$*-ter Ordnung das günstigste.*

Beweis. Eine einfache Auszählung ergibt bei einer algebraischen Gleichung n-ter Ordnung für ein Verfahren (3.8) mit der Ordnung $r \geqq 2$ die Rechenarbeit[1,2]

$$A^* = r(n+1) - 2.$$

Bei m-maliger Durchführung eines Verfahrens r-ter Ordnung haben wir somit die Rechenarbeit

$$A = mA^* = m[r(n+1) - 2]$$

und die Gesamtordnung nach Satz 1

$$k \quad r^m.$$

Also wird

$$A = \frac{\log k}{\log r}[r(n+1) - 2] = Q(r) \log k.$$

Nach Satz 10 ist die günstigste Ordnung r durch diejenige natürliche Zahl $r \geqq 2$ gegeben, für die $Q(r)$ den kleinsten Wert hat. Es ist (r zunächst stetig variabel angenommen), wenn k festgehalten wird,

$$\frac{dA}{dr} = \log k \, \frac{r(n+1)(\log r - 1) + 2}{r(\log r)^2}$$

für $r \geqq 3$ also wegen $\log 3 > 1$ $\frac{dA}{dr} > 0$, d.h. die relative Arbeit (zur Erreichung derselben Gesamtordnung k) nimmt zu, wenn man mit der Ordnung r über 3 hinausgeht. Als günstigste Verfahren kommen daher auch hier wiederum nur die Verfahren der Ordnungen 2 und 3 in Frage. Es ist

$$A(2) = \log k \, \frac{2n}{\log 2} \quad \text{und} \quad A(3) = \log k \, \frac{3n+1}{\log 3}.$$

Daraus folgt

$$A(2) \lessgtr A(3), \quad \text{jenachdem} \quad n \lessgtr \frac{\log 2}{\log \frac{9}{8}} \approx 5{,}885.$$

Damit ist der Satz bewiesen.

Die Sätze 9 und 11 ergeben beide, daß die Verfahren mit einer Ordnung > 3 unter den getroffenen Voraussetzungen unvorteilhaft sind. Ist $F(x)$ kein Polynom, so kann sich jedoch in einzelnen Fällen sehr wohl ein Iterationsverfahren von höherer als 3. Ordnung als das günstigste erweisen. Dies ist insbesondere dann der Fall, wenn sich die Rechenarbeit zur Berechnung der Ableitungen von $F(x)$ wesentlich gegenüber der zur Berechnung von $F(x)$ selbst verringert. Es lohnt sich in vielen Fällen, vor Beginn der Rechnung zu prüfen, welches Verfahren das günstigste ist. Hierzu bietet der folgende Satz eine einfache Möglichkeit:

Satz 12. *Entspricht die Berechnung von* $F(x)$, $F'(x)$, $\frac{F''(x)}{2!}, \ldots, \frac{F^{(r-1)}(x)}{(r-1)!}$ *einer Rechenarbeit von* a_0, $a_1, a_2, \ldots, a_{r-1}$ *Multiplikationen bzw. Divisionen, so ist die Arbeit zur Erzielung einer Gesamtordnung* k *bei Verwendung eines Verfahrens*

[1] Siehe Anmerkung 1, S. 73.

[2] Im Gegensatz zu (4.1) wächst die Rechenarbeit hier nur linear mit der Ordnung k. Dabei wurde vorausgesetzt, daß die Funktionswerte $F_n^{(s)}/s!$ nach dem gewöhnlichen Horner-Schema berechnet werden.

r-ter Ordnung ($r \geqq 2$) *nach* (3.8) *und Fig. 1 gegeben durch*

$$A = \frac{\log k}{2} \cdot \frac{r(r+1)+2(a_0+a_1+\cdots+a_{r-1})-4}{\log r}.$$

Als günstigstes Verfahren erhält man daher dasjenige der Ordnung $r \geqq 2$ *(ganz), bei dem der Quotient*

$$Q = \frac{r(r+1)+2(a_0+a_1+\cdots+a_{r-1})-4}{\log r} \tag{4.4}$$

seinen kleinsten Wert annimmt.

Beweis. Bei einem Schritt mit einem Verfahren der Ordnung r sind nach Satz 8 Gl. (4.1) $M = \frac{r(r+1)-4}{2}$ Multiplikationen bzw. Divisionen durchzuführen. Dazu kommen nach Voraussetzung zur Berechnung von $\frac{F^{(l)}(x)}{l!}$ $(l = 0, 1, 2, \ldots, r-1)$ $a_0+a_1+\cdots+a_{r-1}$ solche Rechenoperationen. Bei n Schritten haben wir somit die Gesamtarbeit

$$A = n \frac{r(r+1)+2(a_0+a_1+\cdots+a_{r-1})-4}{2}$$

und die Gesamtordnung $k = r^n$, also $A = \frac{\log k}{2} Q$.

Alles andere folgt aus Satz 10.

Als Beispiel wollen wir den Fall betrachten, daß die Rechenarbeit zur Berechnung von $F(x)$ an einer Stelle x_n in der Umgebung von ξ a-Multiplikationen entspricht, die zur Berechnung der Ableitungen aber nur $\frac{a}{4}$ Multiplikationen. Dann ergibt sich nach leichter Rechnung als das günstigste Verfahren (3.8)[1]

das Newtonsche Verfahren für	$a < 5{,}019$
das Verfahren 3. Ordnung für	$5{,}019 < a < 20{,}68$
das Verfahren 4. Ordnung für	$20{,}68 < a < 117{,}16$
das Verfahren 5. Ordnung für	$a > 117{,}16$.

§ 5. Fehlerabschätzungen[2]

Aus der Definition der Ordnung ergibt sich, daß für ein Verfahren k-ter Ordnung eine Konstante K_k existiert, so daß

$$\frac{|f(x)-\xi|}{|x-\xi|^k} \leqq K_k \tag{5.1}$$

in der Umgebung der Lösung ξ von $x = f(x)$ bzw. $F(x) = 0$ gilt. Daher muß in jeder Fehlerabschätzung für ein Iterationsverfahren k-ter Ordnung in irgendeiner Weise eine Abschätzung der Konstanten K_k enthalten sein, was im Falle von k-mal differenzierbarem $f(x)$ einer Abschätzung der k-ten Ableitung von $f(x)$

[1] Es sei noch einmal darauf hingewiesen, daß diese Abschätzungen nur für die speziellen nach Satz 7 durchgeführten Verfahren gelten.

[2] L. Collatz bringt in [4] Abschätzungen, die es ermöglichen, für sehr allgemeine Iterationsverfahren beliebiger Ordnung k Fehlerschranken aufzustellen. Hiervon ist bei den obigen spezielleren Fehlerabschätzungen, die sich nur auf die Verfahren des Satzes 7 beziehen, kein Gebrauch gemacht worden.

entspricht, die wiederum im allgemeinen von den k ersten Ableitungen von $F(x)$ abhängt. Hieraus folgt, daß so einfache Fehlerabschätzungen wie sie etwa für lineare Iterationsverfahren ($k=1$) und auch noch für quadratisch konvergente Verfahren möglich sind, für höhere Ordnungen nicht erwartet werden können[1].

Wir betrachten auch hier wieder nur die speziellen nach Satz 7 und Fig. 1 durchgeführten Iterationsverfahren k-ter Ordnung. Ferner sei wieder $F'(\xi) \neq 0$.

Nach (3.2) und (3.3) ist für $k \geqq 3$

$$\xi = x_n - \frac{F_n}{F_n'} - \frac{1}{F_n'}\left\{\sum_{r=2}^{k-1} \frac{(\xi - x_n)^r}{r!} F_n^{(r)} + R_{k-1,n}\right\} \tag{5.2}$$

mit

$$R_{k-1,n} = \int_0^1 \frac{t^{k-1}}{(k-1)!} (\xi - x_n)^k F^{(k)}\big(\xi + t(x_n - \xi)\big)\, dt, \quad \text{also}$$
$$|R_{k-1,n}| \leqq \frac{|\xi - x_n|^k}{k!} \left|F^{(k)}(\xi + \vartheta(x_n - \xi))\right|, \quad 0 < \vartheta < 1. \tag{5.3}$$

Andererseits gilt für unsere Verfahren k-ter Ordnung nach Satz 7

$$x_{n+1} = x_n - \frac{F_n}{F_n'} - \frac{1}{F_n'}\left\{\sum_{r=2}^{k-1} \frac{(\tilde{x}_{n+1} - x_n)^r}{r!} F_n^{(r)}\right\}, \tag{5.4}$$

wobei

$$\tilde{x}_{n+1} = f(x_n)$$

sich aus einem solchen Verfahren $(k-1)$-ter Ordnung berechnet, d.h. es ist

$$\tilde{x}_{n+1} - \xi = g_{k-1}(x_n)\,(x_n - \xi)^{k-1}$$

mit einer in einer Umgebung von ξ beschränkten Funktion $g_{k-1}(x)$. Daher gilt mit $g_{k-1} = g_{k-1}(x_n)$ und $k \geqq 3$:

$$\begin{aligned}(\tilde{x}_{n+1} - x_n)^r &= \left[(\tilde{x}_{n+1} - \xi) + (\xi - x_n)\right]^r \\ &= (\xi - x_n)^r \left[1 - g_{k-1} \cdot (x_n - \xi)^{k-2}\right]^r \\ &= (\xi - x_n)^r + (\xi - x_n)^r \cdot \sum_{\nu=1}^{r} (-1)^\nu \binom{r}{\nu} g_{k-1}^\nu \cdot (x_n - \xi)^{\nu\cdot(k-2)}.\end{aligned}$$

Und hieraus folgt mit (5.2) und (5.4) für $k \geqq 3$:

$$x_{n+1} - \xi = \frac{-1}{F_n'} \sum_{r=2}^{k-1} \left\{\frac{F_n^{(r)}}{r!} \sum_{\nu=1}^{r} (-1)^{r-\nu} \binom{r}{\nu} g_{k-1}^\nu \cdot (x_n - \xi)^{\nu\cdot(k-2)+r}\right\} + \frac{R_{k-1,n}}{F_n'}. \tag{5.5}$$

Wegen $\nu \geqq 1$, $k \geqq 3$, $r \geqq 2$ ist $\nu \cdot (k-2) + r \geqq k$. Damit ergibt (5.5) und (5.3) wiederum

$$x_{n+1} - \xi = O(|x_n - \xi|^k).$$

Die Gl. (5.5) kann nun als Ausgangsgleichung für eine Fehlerabschätzung dienen:

[1] Die Schwierigkeit besteht hier in der Bestimmung einer möglichst kleinen Konstanten K_k in (5.1). Ist diese bekannt, so macht die Fehlerabschätzung keine Mühe.

Nehmen wir zunächst den wichtigsten Fall $k=3$ vorweg, so ergibt (5.5) und (5.3)

$$x_{n+1}-\xi=\frac{-1}{F_n'}\frac{F_n''}{2}\left[-2g_2(x_n-\xi)^3+g_2^2(x_n-\xi)^4\right]+\frac{R_{2,n}}{F_n'} \tag{5.6}$$
$$=\frac{(\xi-x_n)^3}{2F_n'}\left[-2F_n''g_2+F_n''g_2^2(x_n-\xi)+\frac{1}{3}\tilde{F}'''\right],$$

wobei $\tilde{F}'''$ bedeutet, daß die Funktion $F'''(x)$ an einer Zwischenstelle zu nehmen ist.

Aus (3.9) folgt

$$g_2=\frac{-\tilde{F}''}{2F_n'},$$

so daß sich

$$x_{n+1}-\xi=\frac{(\xi-x_n)^3}{2F_n'}\left[\frac{F_n''\tilde{F}''}{F_n'}+\frac{F_n''\tilde{F}''^2}{4F_n'^2}(x_n-\xi)+\frac{1}{3}\tilde{F}'''\right] \tag{5.7}$$

ergibt. Aus dieser letzten Gleichung folgt nun leicht folgende Fehlerabschätzung für das Verfahren 3. Ordnung:

Satz 13. *Es existiere ein Bereich $\mathfrak{B}$, in dem folgendes gilt:*

α) $$|F'(x)|\geqq m>0 \quad \textit{und} \quad |F^{(l)}(x)|\leqq M_l$$

mit Konstanten m und M_l, $l=2,3$.

β) $$\frac{|F(x_0)|}{m}\leqq c$$

mit einer Konstanten c.

γ) *Es sei*

$$Kc^2<1 \quad \textit{mit} \quad K=\frac{1}{2m^3}\left[mM_2^2+\frac{1}{4}M_2^3c+\frac{1}{3}m^2M_3\right].$$

Liegen dann alle x mit

$$|x-x_0|\leqq 2c \tag{5.8}$$

in $\mathfrak{B}$, so konvergiert das Iterationsverfahren (3.8) (Satz 7) für $k=3$ gegen die einzige Lösung ξ von $F(x)=0$ in $\mathfrak{B}$ und es gilt die Fehlerabschätzung

$$|x_n-\xi|\leqq K^{\frac{1}{2}(3^n-1)}c^{3^n}, \qquad n=0,1,2,\ldots. \tag{5.9}$$

Beweis. Wegen α) und β) existiert genau eine Nullstelle ξ von $F(x)$ in $\mathfrak{B}$, und es ist sogar

$$|\xi-x_0|\leqq c. \tag{5.10}$$

Es gelte $|x_n-\xi|\leqq c$ für $n=0,1,\ldots,r$. Dann liegen wegen

$$|x_n-x_0|\leqq|x_0-\xi|+|x_n-\xi|\leqq 2c$$

diese x_n in $\mathfrak{B}$, und es folgt aus (5.7), α) und γ)

$$|x_{r+1}-\xi|\leqq|x_r-\xi|^3\frac{1}{2m}\left[\frac{M_2^2}{m}+\frac{M_2^3}{4m^2}c+\frac{M_3}{3}\right]=K|x_r-\xi|^3$$
$$\leqq Kc^3<c.$$

Daher liegen alle x_n in dem Bereich $|x-\xi|\leqq c$ von $\mathfrak{B}$ und es gilt

$$|x_{n+1}-\xi|\leqq K|x_n-\xi|^3, \qquad n=0,1,2,\ldots. \tag{5.11}$$

Wegen $K|x_n-\xi|^2\leqq Kc^2<1$ konvergiert das Iterationsverfahren, und durch mehrmalige Anwendung von (5.11) ergibt sich mit (5.10) die Fehlerabschätzung (5.9), wie man auch leicht durch vollständige Induktion nachprüft, w.z.b.w.

Wegen $F'(\xi)\neq 0$ läßt sich offenbar stets ein solcher Bereich $\mathfrak{B}$ angeben, der die geforderten Bedingungen erfüllt, wenn nur x_0 nahe genug bei ξ liegt[1]. Daher ist die Fehlerabschätzung immer durchführbar.

Für die Verfahren (3.8) der Ordnung $k>3$ wird die Bestimmung der Konstanten K_k in (5.1) etwas mühsamer.

Aus (5.5) und (5.3) folgt wegen $\nu\cdot(k-2)+r\geqq\nu+k-1$ für $\nu\geqq 1$, $r\geqq 2$ und $k\geqq 3$, wenn wir bereits wissen, daß $|x_n-\xi|\leqq 1$ ist[2],

$$\begin{aligned}|x_{n+1}-\xi|&\leqq\frac{|x_n-\xi|^{k-1}}{|F_n'|}\left\{\sum_{r=2}^{k-1}\frac{|F_n^{(r)}|}{r!}\sum_{\nu=1}^{r}\binom{r}{\nu}|g_{k-1}^{\nu}|\,|x_n-\xi|^{\nu}\right\}+\frac{|\tilde F^{(k)}|}{k!\,|F_n'|}\cdot|x_n-\xi|^k\\&\leqq\frac{|x_n-\xi|^{k-1}}{|F_n'|}\left[(1+|g_{k-1}|\,|x_n-\xi|)^{k-1}-1\right]\cdot\sum_{r=2}^{k-1}\frac{|F_n^{(r)}|}{r!}+\frac{|\tilde F^{(k)}|}{k!\,|F_n'|}|x_n-\xi|^k.\end{aligned} \tag{5.12}$$

Dabei bedeutet $\tilde F^{(k)}$, daß der Funktionswert von $F^{(k)}(x)$ an einer Zwischenstelle zwischen x_n und ξ zu nehmen ist. Die Summe

$$S_{k-1}=\sum_{r=2}^{k-1}\frac{|F_n^{(r)}|}{r!}=\frac{|F_n''|}{2!}+\frac{|F_n'''|}{3!}+\cdots+\frac{|F_n^{(k-1)}|}{(k-1)!}$$

kann unmittelbar aus dem Schema (Fig. 1) entnommen werden. Die Gl. (5.12) liefert eine Rekursionsformel zur Abschätzung der Größen $g_l=g_l(x_n)$:

$$|g_l|\leqq\frac{1}{|F_n'|}\left\{\frac{S_{l-1}}{|x_n-\xi|}\left[(1+|g_{l-1}|\,|x_n-\xi|)^{l-1}-1\right]+\frac{|\tilde F^{(l)}|}{l!}\right\},\quad l\geqq 3. \tag{5.13}$$

Hieraus ergeben sich mit

$$|F'(x)|\geqq m>0,\qquad |F^{(l)}(x)|\leqq M_l,\qquad l=2,3,\ldots,k,\qquad x\in\mathfrak{B} \tag{5.14}$$

die Konstanten K_k in (5.1) rekursiv aus der Formel

$$K_l=\frac{1}{m}\left\{\frac{S_{l-1}^*}{|x_n-\xi|}\left[(1+K_{l-1}|x_n-\xi|)^{l-1}-1\right]+\frac{M_l}{l!}\right\},\quad l\geqq 3, \tag{5.15}$$

wobei $S_{l-1}^*=\sum_{r=2}^{l-1}\frac{M_r}{r!}$ und $K_2=\frac{M_2}{2m}$ ist.

Analog zu Satz 13 ergibt sich damit der

Satz 14. *Gelten in einem Bereich $\mathfrak{B}$ die Abschätzungen* (5.14), *ist* $\frac{|F(x_0)|}{m}\leqq c<1$, *liegen ferner alle x mit $|x-x_0|\leqq 2c$ in $\mathfrak{B}$ und gilt $K_k c^{k-1}<1$, wobei sich K_k*

[1] In diesem Falle muß jedoch die Existenz der Lösung ξ von $F(x)=0$ vorausgesetzt werden.

[2] Dies ist offenbar der Fall, wenn $|F'(x)|\geqq m>0$ und $|F(x_n)|\leqq cm<m$ in $|x-x_n|\leqq c$ gilt.

rekursiv aus den Formeln

$$(5.16)\quad K_2 = \frac{M_2}{2m}, \quad K_l = \frac{1}{m}\left\{\frac{S^*_{l-1}}{c}\left[(1+K_{l-1}c)^{l-1}-1\right]+\frac{M_l}{l!}\right\}, \quad 3 \leqq l \leqq k,$$

berechnet, so konvergiert das Verfahren (3.8) *k-ter Ordnung gegen die einzige Lösung* ξ *von* $F(x)=0$ *in* $\mathfrak{B}$, *und es gilt die Fehlerabschätzung*

$$(5.17)\qquad |x_n-\xi| \leqq K_k^{\frac{k^n-1}{k-1}} c^{k^n}, \qquad n=0,1,2,\ldots.$$

Der *Beweis* ist völlig analog dem von Satz 13. Auch hier existiert wegen $F'(\xi)\neq 0$ stets ein Bereich $\mathfrak{B}$, der die geforderten Bedingungen erfüllt. Wegen $c<1$ kann man die 2. Gleichung in (5.16) auch durch die gröbere Rekursionsformel

$$K_l = \frac{1}{m}\left\{S^*_{l-1}\left[(1+K_{l-1})^{l-1}-1\right]+\frac{M_l}{l!}\right\}, \quad 3 \leqq l \leqq k,$$

ersetzen.

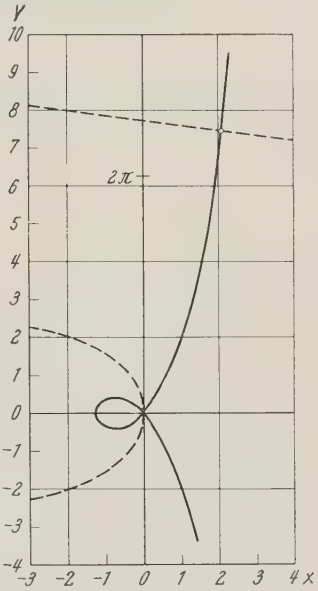

Fig. 2. Erläuterung zum 1. Beispiel

§ 6. Beispiele

1. Beispiel. $e^z=1+z$, $z=x+yi\neq 0$.

α) Als Näherungswert für eine Lösung ζ ergibt sich aus der Zeichnung (Fig. 2) als Schnitt der Kurven gleichen Betrages und gleichen Winkels[1] der Wert $z_0=2{,}1+7{,}5\,i$.

β) Entspricht die Berechnung von $F_n=e^{z_n}-z_n-1$ und $-1/F_n'$ der Rechenarbeit von $a=a_0+a_1$ Multiplikationen (komplexer Zahlen) und vernachlässigt man dabei die Arbeit zur Berechnung der Zahlen[2] $F_n^{(s)}/s!$ $(s\geqq 2)$, so ergibt sich nach Satz 12 als die günstigste Ordnung r diejenige, bei der der Quotient (4.4)

$$(6.1)\qquad Q=\frac{r(r+1)+2a-4}{\log r}$$

seinen kleinsten Wert[3] annimmt. Es ist also das Verfahren der Ordnung $k=r$ $(r\geqq 2)$ günstiger (ungünstiger) als das Verfahren der Ordnung $k=r+1$, wenn $Q(r)$ kleiner (größer) als $Q(r+1)$ ist, woraus nach (6.1)

$$a \underset{(>)}{\leq} \frac{1}{2}\,\frac{[(r+1)(r+2)-4]\log r-[r(r+1)-4]\log(r+1)}{\log(r+1)-\log r}=G(r)$$

folgt. Es ergibt[4] sich für

$r=$	2	3	4	5	6	7	8	9
$G(r)=$	4,129	11,275	23,063	39,966	62,360	90,586	124,902	165,526

[1] Beide Kurven lassen sich leicht punktweise mit Hilfe der Kreis- bzw. Geradenschar durch den Punkt $(-1, 0)$ konstruieren. Sie sind symmetrisch zur x-Achse.

[2] Diese lassen sich etwa für $s\leqq 10$ unmittelbar fortlaufend hinschreiben. Wesentlich zur Berechnung der Funktionswerte $F_n^{(s)}/s!$ ist hier in bezug auf die Rechenarbeit nur die Berechnung von

$$e^{z_n}=e^{x_n}(\cos y_n+i\sin y_n).$$

[3] Die Basis des Logarithmus ist dabei beliebig >1.

[4] In der Praxis würde natürlich hier eine viel gröbere Schätzung genügen.

Ist also z.B. $a<4{,}129$, so ist das Verfahren 2. Ordnung günstiger als das 3. Ordnung; ist $a<11{,}275$, so ist das Verfahren 3. Ordnung günstiger als das 4. Ordnung usw.

Die Wahl der Ordnung hängt hier wesentlich davon ab, welche Genauigkeit erzielt werden soll. Für größere Genauigkeiten werden die üblichen Tafeln im allgemeinen nicht mehr ausreichen, und man hat die Funktionswerte (insbesondere e^{z_n}) selbst zu berechnen. Daher wird man in diesem Fall zweckmäßig ein Verfahren höherer Ordnung wählen[1].

Tabelle 1. *Iterationsverfahren k-ter Ordnung* $z_1=z_0+v_k$ *zur Bestimmung*

			$\frac{F''(z_0)}{2!}=$
			1,41534138+3,82993368i
		$\frac{F'''(z_0)}{3!}=$	0,04390237−0,03244396i
		0,47178046+1,27664456i	1,45924375+3,79748972i
	$\frac{F^{(4)}(z_0)}{4!}=$	0,01097556−0,00810316i	0,04346772−0,03274491i
	0,11794512+0,31916114i	0,48275602+1,26854140i	1,45880910+3,79718877i
$\frac{F^{(5)}(z_0)}{5!}=$	0,00219505−0,00162060i	0,01088833−0,00816948i	0,04346482−0,03274027i
0,02358902+0,06383223i	0,12014017+0,31754054i	0,48266879+1,26847508i	1,45880620+3,79719341i

γ) Wir führen hier die Verfahren bis zur 6. Ordnung durch (Tabelle 1). Dabei wurde 8-stellig gerechnet, um Abrundungsfehler zu vermeiden. Diese können sich hier nicht aufschaukeln, wie man sich leicht an Hand der Rechnung überzeugt. Geht man mit der Ordnung k über 6 hinaus, so ändern sich die v_k nicht mehr innerhalb der Genauigkeit von 8 Stellen nach dem Komma:

$$v_k = v_6 \pm 0{,}5 \cdot 10^{-8}.$$

Mit $z_1=z_0+v_k$ und $z_0=2{,}1+7{,}5\,i$ erhalten wir somit die folgenden Ergebnisse bei einem Schritt mit dem

Verfahren 2. Ordnung (NEWTON): $z_1 = 2{,}0882059 + 7{,}4620217\,i$
Verfahren 3. Ordnung: $z_1 = 2{,}0882215 + 7{,}4614801\,i$
Verfahren 4. Ordnung: $z_1 = 2{,}0884330 + 7{,}4614882\,i$
Verfahren 5. Ordnung: $z_1 = 2{,}0884330 + 7{,}4614892\,i$
Verfahren 6. Ordnung: $z_1 = 2{,}0884330 + 7{,}4614893\,i$.

[1] Wegen der einfachen geschlossenen Programmierung der Verfahren gilt dies auch bei Benutzung einer elektronischen Rechenanlage.

δ) Eine grobe Schätzung läßt erwarten, daß für den Konvergenznachweis und die Fehlerabschätzung nach Satz 14 der Bereich

$$\mathfrak{B}: |z - z_0| \leqq 0{,}1 \tag{6.2}$$

die geforderten Bedingungen erfüllt. Aus (6.2) folgt mit $z = x + iy$

$$\left.\begin{array}{l} 2{,}0 \leqq x \leqq 2{,}2 \\ 7{,}4 \leqq y \leqq 7{,}6 \end{array}\right\} \quad \text{und daraus also} \quad |F'(z)| = |e^z - 1| \geqq |e^z| - 1 = e^x - 1,$$
$$|F'(z)| \geqq e^2 - 1 = 6{,}389 = m, \qquad z \in \mathfrak{B}.$$

einer Lösung der Gleichung $F(z) = e^z - z - 1 = 0$, $(z_0 = 2{,}1 + 7{,}5i)$

	$F(z_0) =$ $-0{,}26931725 + 0{,}15986737\,i$	$\frac{-1}{F'(z_0)} =$ $-0{,}02951526 + 0{,}12349652\,i$
		$v_2 = -0{,}01179410 - 0{,}03797827\,i$
0		
	$-0{,}00527555 - 0{,}00372344\,i$	
$0{,}12876158 - 0{,}09892284\,i$	$-0{,}27459280 + 0{,}15614393\,i$	$v_3 = -0{,}01117855 - 0{,}03851988\,i$
	$-0{,}00525322 - 0{,}00390342\,i$	
$0{,}12996662 - 0{,}09866032\,i$	$-0{,}27457047 + 0{,}15596395\,i$	$v_4 = -0{,}01115699 - 0{,}03851181\,i$
	$-0{,}00524518 - 0{,}00390554\,i$	
$0{,}12996069 - 0{,}09854658\,i$	$-0{,}27456243 + 0{,}15596183\,i$	$v_5 = -0{,}01115696 - 0{,}03851076\,i$
	$-0{,}00524496 - 0{,}00390528\,i$	
$0{,}12995696 - 0{,}09854487\,i$	$-0{,}27456221 + 0{,}15596209\,i$	$v_6 = -0{,}01115700 - 0{,}03851074\,i$

Weiter wird

$$\frac{|F(z_0)|}{m} = 0{,}0490 < 0{,}05 = c < 1,$$

und damit gilt $|z - z_0| \leqq 2c$ in $\mathfrak{B}$.

Ferner ergeben sich unmittelbar die Abschätzungen

$$|F^{(l)}(z)| = |e^z| = e^x \leqq e^{2{,}2} = 9{,}0250135 \leqq 9{,}0251 = M_l, \qquad l = 2, 3, \ldots$$

in $\mathfrak{B}$, und hieraus rekursiv aus der Formel (5.16) folgende Konstanten K_k bei den Verfahren der Ordnung:

(6.3)

$k =$	2	3	4	5	6
$K_k =$	0,70631	1,2508	3,8183	20,2544	648,556

Es folgt $K_k c^{k-1} < 1$ für $k = 2, 3, \ldots, 6$. Daher konvergieren diese Verfahren.

Unter Verwendung der Konstanten K_k ergeben sich jeweils nach dem n-ten Schritt aus der Fehlerformel (5.17) die Schranken für das Verfahren

2. Ordnung: $|z_n - \zeta| \leqq 1{,}42\,(3{,}532 \cdot 10^{-2})^{2^n}$
3. Ordnung: $|z_n - \zeta| \leqq 0{,}895\,(5{,}592 \cdot 10^{-2})^{3^n}$
4. Ordnung: $|z_n - \zeta| \leqq 0{,}639\,(7{,}830 \cdot 10^{-2})^{4^n}$
5. Ordnung: $|z_n - \zeta| \leqq 0{,}472\,(1{,}0608 \cdot 10^{-1})^{5^n}$
6. Ordnung: $|z_n - \zeta| \leqq 0{,}274\,(1{,}8255 \cdot 10^{-1})^{6^n}$.

Noch genauere Fehlerschranken erhält man, wenn man die Fehlerabschätzungen von Satz 14 erst *nach* dem ersten Schritt durchführt. Für den Wert $z_1 = 2{,}088\,843\,0 + 7{,}461\,489\,3\,i$ des Verfahrens 6. Ordnung ergibt sich dabei unter Verwendung des neuen Bereiches $\mathfrak{B}_1$, $|z - z_1| \leqq 10^{-7}$, ein Fehler von $|z_1 - \zeta| < 0{,}15 \cdot 10^{-7}$. z_1 ist also noch in den letzten Stellen genau.

2. Beispiel. $F(x) \equiv x^7 + 5x^6 + 3x^5 + 2x^4 + 4x^3 + 2x^2 + 6x + 1 = 0$.

Es ist $F(-1) = -1$ und $F(-0{,}5) = +1{,}1016$. Daher liegt zwischen $x = -1$ und $x = -0{,}5$ eine reelle Nullstelle von $F(x)$. Wir wählen $x_0 = -0{,}75$.

Nach Satz 11 ist hier das Verfahren 3. Ordnung das günstigste.

Zur Veranschaulichung der Verfahren wollen wir hier trotzdem das Schema der Fig. 1 über $k = 3$ hinaus weiterrechnen. Wir erhalten so die Tabelle 2 und damit das Ergebnis nach einem Schritt bei dem

Verfahren 2. Ordnung (NEWTON):	$x_1^{(2)} = -0{,}676\,599\,2$
Verfahren 3. Ordnung:	$x_1^{(3)} = -0{,}682\,595\,7$
Verfahren 4. Ordnung:	$x_1^{(4)} = -0{,}680\,708\,1$
Verfahren 5. Ordnung:	$x_1^{(5)} = -0{,}680\,993\,5$
Verfahren 6. Ordnung:	$x_1^{(6)} = -0{,}680\,958\,6$
Verfahren 7. Ordnung:	$x_1^{(7)} = -0{,}680\,962\,6$
Verfahren 8. Ordnung:	$x_1^{(8)} = -0{,}680\,962\,2$.

Bei Weiterrechnung des Schemas ergibt sich durch die Verfahren 9. und höherer Ordnung innerhalb der Rechengenauigkeit von 7 Stellen keine Änderung mehr.

Führen wir statt dessen zweimal das (günstigste) Verfahren 3. Ordnung durch, so erhalten wir (mit $x_0 = -0{,}75$; $x_1 = -0{,}682\,595\,7$) den Wert $x_2 = -0{,}680\,962\,2$, also denselben wie bei einem Schritt mit dem Verfahren 8. Ordnung.

Es ist $F(-0{,}680\,962\,2) = 4{,}4 \cdot 10^{-8}$, also $|F(x_1^{(8)})| < 5 \cdot 10^{-8}$. In dem Bereich $\mathfrak{B}$: $-0{,}681 \leqq x \leqq -0{,}680$ ist $|F'(x)| > 5{,}8 = m$. Es folgt

$$\frac{|F(x_1^{(8)})|}{m} \leqq \frac{5}{5{,}8}\, 10^{-8} < 1 \cdot 10^{-8} = c.$$

Damit haben wir das Ergebnis

$$\underline{\xi = -0{,}680\,962\,20 \pm 10^{-8}.}$$

Für die Größen $\frac{F^{(l)}(x)}{l!}$ ergeben sich in dem Bereich $\mathfrak{B}$ die oberen Schranken $\frac{M_2}{2!} = 3$, $\frac{M_3}{3!} = 12$, $\frac{M_4}{4!} = 16$, $\frac{M_5}{5!} = 8$ $\left(F^{(l)}(x) \leqq M_l,\ x \in \mathfrak{B}\right)$. Damit erhält man (mit $c = 10^{-8}$) nach leichter Rechnung aus (5.16) und (5.17)

die Konstanten[1]	und die Fehlerabschätzungen
$K_2 = 0{,}5173$	$\lvert x_n - \xi\rvert \leqq 1{,}93\,[0{,}52 \cdot 10^{-8}]^{2^n}$
$K_3 = 2{,}61$	$\lvert x_n - \xi\rvert \leqq 0{,}62\,[1{,}62 \cdot 10^{-8}]^{3^n}$
$K_4 = 23{,}1$	$\lvert x_n - \xi\rvert \leqq 0{,}36\,[2{,}85 \cdot 10^{-8}]^{4^n}$
$K_5 = 493{,}9$	$\lvert x_n - \xi\rvert \leqq 0{,}22\,[4{,}72 \cdot 10^{-8}]^{5^n}$

für die Iterationsverfahren 2. bis 5. Ordnung, wobei $x_0 = -0{,}680\,962\,20$ gesetzt ist.

[1] Die verhältnismäßig großen Beträge der Ableitungen $F^{(l)}(x)$ bei diesem Beispiel wirken sich auf die Konvergenz und die Abschätzungen bei höheren Verfahren ungünstig aus.

Tabelle 2. *Schema zur Bestimmung einer Nullstelle der Gleichung*

$x^7+5x^6+3x^5+2x^4+4x^3+2x^2+6x+4=0$

durch ein Iterationsverfahren k-ter Ordnung: $x_1=x_0+v_k$. $x_0=-0{,}75$

1	5	3	2	4	2	6	4	
	−0,75	−3,1875	0,140625	−1,6054688	−1,7958985	−0,1530761	−4,3851929	
1	4,25	−0,1875	2,140625	2,3945312	0,2041015	5,8469239	−0,3851929	$v_2=0{,}0734008$
	−0,75	−2,625	2,109375	−3,1875000	0,5947266	−0,5991211		
1	3,50	−2,8125	4,250000	−0,7929688	0,7988281	(5,2478028)		
	−0,75	−2,0625	3,656250	−5,9296875	5,0419922		0,0314685	
1	2,75	−4,8750	7,90625	−6,7226563	5,8408203	0,4287209	−0,3537244	$v_3=0{,}0674043$
	−0,75	−1,5	4,78125	−9,515625	−1,0948300		0,0215626	
1	2	−6,375	12,68750	−16,2382813	4,7459903	0,3199002	−0,3636303	$v_4=0{,}0692919$
	−0,75	−0,9375	5,484375	1,2591637	−1,0379315		0,0230604	
1	1,25	−7,3125	18,1718750	−14,9791176	4,8028888	0,3328013	−0,3621325	$v_5=0{,}0690065$
	−0,75	−0,375	−0,5304875	1,2173704	−1,0365405		0,0228775	
1	0,5	−7,6875	17,6413875	−15,0209109	4,8042798	0,3315265	−0,3623154	$v_6=0{,}0690414$
	−0,75	−0,0172604	−0,5319474	1,2178853	−1,0370292		0,0228983	
1	−0,25	−7,7047604	17,6399276	−15,0203960	4,8037911	0,3316605	−0,3622946	$v_7=0{,}0690374$
	0,0690374	−0,0124932	−0,5315875	1,2178396	−1,0369722		0,0228959	
1	−0,1809626	−7,6999932	17,6402875	−15,0204417	4,8038481	0,3316452	−0,3622970	$v_8=0{,}0690378$
	0,0690378	−0,0124932	−0,5315906	1,2178464	−1,0369778		0,0228962	
1	−0,1809622	−7,6999932	17,6402844	−15,0204349	4,8038425	0,3316467	−0,3622967	$v_9=0{,}0690378=v_8$

wie Verfahren 9. Ordnung

Literatur

[1] Schröder, E.: Über unendlich viele Algorithmen zur Auflösung der Gleichungen. Math. Ann. **2**, 317—365 (1870).
[2] Bodewig, E.: Konvergenztypen und das Verhalten von Approximationen in der Nähe einer mehrfachen Wurzel einer Gleichung. Z. angew. Math. Mech. **29**, 44—51 (1949).
[3] Ludwig, R.: Über Iterationsverfahren für Gleichungen und Gleichungssysteme. Z. angew. Math. Mech. **34**, 210—225 (1954).
[4] Collatz, L.: Näherungsverfahren höherer Ordnung für Gleichungen in Banach-Räumen. Arch. Rational Mech. Anal. **2**, 66—75 (1958).
[5] Zurmühl, R.: Praktische Mathematik. Berlin-Göttingen-Heidelberg: Springer 1957.
[6] Hartree, D. R.: Notes on iterative processes. Proc. Cambridge phil. Soc. **45**, 230—236 (1949).
[7] Domb, C.: On iterative solutions of algebraic equations. Proc. Cambridge phil. Soc. **45**, 237—240 (1949).
[8] Kiss, I.: Über eine Verallgemeinerung des Newtonschen Näherungsverfahrens. Z. angew. Math. Mech. **34**, 68—69 (1954).
[9] Zajta, A.: Untersuchungen über die Verallgemeinerung der Newton-Raphsonschen Wurzelapproximation. Acta techn. acad. scient. Hungaricae **15**, 233—260 (1956).

Mathematisches Institut A
der Technischen Hochschule
Stuttgart

(Eingegangen am 17. Juli 1959)

A Characterization of Hypergeometric Functions

N. Chako & J. Meixner

Summary

A theory of special functions should derive as many of the numerous formulas and properties as possible from some guiding principles. One of these principles is the F-equation, which Truesdell has put forward, and which is satisfied by all simple special functions. Many results about addition theorems, generating functions, integral representations *etc.* follow immediately from the F-equation. One disadvantage, however, is that the F-equation does not constitute a specific property of the simple special functions, generating instead functions of a much more general character.

By assuming two functional relations one of which can be transformed into an ascending and the other into a descending F-equation, one obtains as solutions of these functional relations the hypergeometric functions with all their special and confluent cases, and thus one has a new characterization of the hypergeometric functions.

1. Introduction

By the simple special functions of mathematical physics we understand the hypergeometric functions with their special and limiting cases, examples being Legendre functions, Bessel functions and confluent hypergeometric functions.

A common property of these simple special functions is that if they are written in appropriate variables, they become solutions of linear homogeneous differential equations of second order with polynomial coefficients. But this is a property common also to other special functions, such as Mathieu and spheroidal functions.

Another common property is the existence of linear three-term recurrence relations between three contiguous functions, where contiguous functions are defined as those differing by one in some parameter. The coefficients in these recurrence relations are polynomials in the independent variable and in the parameters, if both are suitably defined. But without specifying further the nature of the coefficients, these recurrence relations do not characterize of the simple special functions; such relations also hold true, for instance, for the Fourier expansion coefficients of Mathieu functions.

Truesdell* has drawn attention to the fact that the simple special functions, after appropriate transformation, satisfy a functional equation of the form

$$\frac{dF(z,\alpha)}{dz} = F(z,\alpha+1) \qquad (\alpha=\alpha_0,\ \alpha_0\pm 1,\ \alpha_0\pm 2,\ldots), \tag{1}$$

* Truesdell, C.: An Essay toward a Unified Theory of Special Functions. Princeton University Press 1948.

which is called TRUESDELL's F-equation. This property (1) is remarkable and powerful because it permits one to derive in a brief and straightforward fashion many theorems for simple special functions; in particular, addition theorems, generating functions, integral relations. But this F-equation is not a characterizing property of the simple special functions either, since the generalized zeta function $\zeta(\alpha, z)$, after multiplication by $e^{i\alpha\pi}\Gamma(\alpha)$, satisfies (1) without belonging to the functions of hypergeometric type.

However, the power of functional equations like (1) can be applied to a characterization of the simple special functions, if one recalls that the simple special functions always satisfy two simultaneous functional equations of the types

$$\frac{dy(z,\alpha)}{dz} = U_1(z,\alpha)\, y(z,\alpha) + V_1(z,\alpha)\, y(z,\alpha+1), \tag{2}$$

$$\frac{dy(z,\alpha)}{dz} = U_2(z,\alpha)\, y(z,\alpha) + V_2(z,\alpha)\, y(z,\alpha-1). \tag{3}$$

By themselves, these equations do not give a sufficient restriction on the possible solutions of (2) and (3), so long as we do not stipulate additional conditions. We require therefore, that there exist functions $z(\xi_i)$ and $\Phi_i(z(\xi_i), \alpha)$, $i=1,2$, such that the functions $Y_i(\xi_i, \alpha) = \Phi_i(z(\xi_i), \alpha)\; y(z(\xi_i), \alpha)$ satisfy the ascending F-equation

$$\frac{dY_1(\xi_1,\alpha)}{d\xi_1} = Y_1(\xi_1, \alpha+1)$$

and the descending F-equation

$$\frac{dY_2(\xi_2,\alpha)}{d\xi_2} = Y_2(\xi_2, \alpha-1).$$

Even so, too much arbitrariness remains, and therefore we require in addition that there be two linearly independent solutions of (2) and (3) with the above mentioned properties.

There is no loss in generality if we assume $U_1(z,\alpha)=0$, $V_1(z,\alpha)=1$, which simply means that we transform the equations (2) and (3) by taking $Y_1(\xi_1,\alpha)$ as new dependent variable.

Changing the notation slightly, we therefore seek functions $y(z,\alpha)$, $\alpha=\alpha_0$, $\alpha_0\pm1$, $\alpha_0\pm2$, ..., with the following properties*:

1. $y(z,\alpha)$ is a solution of the ascending F-equation

$$\frac{dy(z,\alpha)}{dz} = y(z,\alpha+1). \tag{4}$$

2. There exists a function $\Phi(z,\alpha)$ and a variable $\xi=\xi(z)$ with $d\xi/dz \neq 0$ in some domain of the complex variable z such that

$$y(\xi,\alpha) = \Phi(z(\xi),\alpha)\, y(z(\xi),\alpha) \tag{5}$$

is a solution of the descending F-equation

$$\frac{dY(\xi,\alpha)}{d\xi} = Y(\xi,\alpha-1). \tag{6}$$

3. There exists a second function $y_1(z,\alpha)$, which is linearly independent of $y(z,\alpha)$ and has the same properties 1. and 2.

* This is a generalization of a problem which TRUESDELL indicated in Theorem 6.1 of his monograph.

2. Derivation of the general solution

We note first the following

Lemma. *If $y(z, \alpha)$ is a solution of the ascending F-equation, so also is*

$$k^{\alpha} y(kz+l, \alpha+a).$$

If $Y(\xi, \alpha)$ is a solution of the descending F-equation, so also is

$$k^{-\alpha} Y(k\xi+l, \alpha+a).$$

Here $k, l,$ and a are arbitrary constants with the restriction $k \neq 0$.

The proof of this lemma is elementary.

In the following considerations, for simplicity of notation, we often omit the variable z from the arguments of the functions $y(z, \alpha)$ and $\Phi(z, \alpha)$. A prime means differentiation with respect to the variable z.

By inserting (5) into (6), we obtain the equation

$$[\Phi'(\alpha)\, y(\alpha) + \Phi(\alpha)\, y'(\alpha)] \frac{dz}{d\xi} = \Phi(\alpha-1)\, y(\alpha-1). \tag{7}$$

The equations (7) and (4) are not compatible unless some conditions on the functions $\Phi(\alpha)$ and $z(\xi)$ hold. To derive them, we replace α by $\alpha-1$ in (4) and by $\alpha+1$ in (7). Let us denote the resulting equations by (4′) and (7′). Then we eliminate $y(\alpha-1)$ from (4′) and (7) and $y(\alpha+1)$ from (4) and (7′). This elimination results in two linear and homogeneous differential equations of second order for $y(z, \alpha)$. They are

$$\begin{aligned}&[\Phi''(\alpha)\, y(\alpha) + 2\Phi'(\alpha)\, y'(\alpha) + \Phi(\alpha)\, y''(\alpha)] \frac{dz}{d\xi} + \\ &+ [\Phi'(\alpha)\, y(\alpha) + \Phi(\alpha)\, y'(\alpha)] \left[\frac{d}{dz}\left(\frac{dz}{d\xi}\right) - \frac{\Phi'(\alpha-1)}{\Phi(\alpha-1)} \frac{dz}{d\xi}\right] = \Phi(\alpha-1)\, y(\alpha),\end{aligned} \tag{8}$$

$$[\Phi(\alpha+1)\, y''(\alpha) + \Phi'(\alpha+1)\, y'(\alpha)] \frac{dz}{d\xi} - \Phi(\alpha)\, y(\alpha) = 0. \tag{9}$$

Since both differential equations are satisfied by the same function $y(z, \alpha)$ and by the same linearly independent function $y_1(z, \alpha)$, they must be identical. By comparing their coefficients, one therefore obtains the following two equations for $\Phi(\alpha)$ and $z(\xi)$:

$$\frac{\Phi'(\alpha+1)}{\Phi(\alpha+1)} - 2\frac{\Phi'(\alpha)}{\Phi(\alpha)} + \frac{\Phi'(\alpha-1)}{\Phi(\alpha-1)} = \frac{(dz/d\xi)'}{dz/d\xi}, \tag{10}$$

$$\frac{\Phi''(\alpha)}{\Phi(\alpha)} \frac{dz}{d\xi} + \frac{\Phi'(\alpha)}{\Phi(\alpha)} \frac{d}{dz}\left(\frac{dz}{d\xi}\right) - \frac{\Phi'(\alpha)\,\Phi'(\alpha-1)}{\Phi(\alpha)\,\Phi(\alpha-1)} \frac{dz}{d\xi} + \frac{\Phi(\alpha)}{\Phi(\alpha+1)} - \frac{\Phi(\alpha-1)}{\Phi(\alpha)} = 0. \tag{11}$$

The first equation constitutes a linear difference equation in α of second order for the function $\Phi'(\alpha)/\Phi(\alpha)$, which is solved by a polynomial in α of second degree,

$$\frac{\Phi'(\alpha)}{\Phi(\alpha)} = p_0'(z) + \alpha\, p_1'(z) + \alpha^2 p_2'(z), \tag{12}$$

with coefficients $p_0'(z)$, $p_1'(z)$, $p_2'(z)$ which may depend on z. Now the integration with respect to z can be performed. It gives

$$\Phi(\alpha) = q(\alpha) \exp[p_0(z) + \alpha\, p_1(z) + \alpha^2 p_2(z)]. \tag{13}$$

The factor $q(\alpha)$ is the integration constant and thus depends only on α.

Due to (12), the first three terms in (11) form a polynomial of at most fourth degree in α. Considering (11) as a difference equation for $\Phi(\alpha)/\Phi(\alpha+1)$, one observes that this expression has to be a polynomial of fifth degree or less in α. On the other hand, one obtains from (13)

$$\frac{\Phi(\alpha)}{\Phi(\alpha+1)} = \frac{q(\alpha)}{q(\alpha+1)} \exp[-p_1(z) - (2\alpha+1)\, p_2(z)], \tag{14}$$

which is a polynomial in α for all z, only if $p_2(z)$ is a constant. Without loss of generality, one can put this constant equal to zero by redefining $q(\alpha)$ as the $q(\alpha)$ in (13) multiplied by $\exp(\alpha^2 p_2)$.

Thus, putting $p_2(z)=0$, one concludes from (12) that $\Phi'(\alpha)/\Phi(\alpha)$ is a polynomial of degree at most one in α, and therefore the right member of (10) vanishes. This is equivalent to z's being a linear function of ξ and vice versa. By the lemma, one can restrict consideration to the special case $z=\xi$ and then arrive at the general function $Y(\xi, \alpha)$ connected with $y(z, \alpha)$ by the transformation given in the lemma.

What remains now is the discussion of equation (11). Making use of (13), of $p_2=0$ and $z=\xi$, we simplify (11) to the form

$$\begin{aligned} &p_0''(z) + p_0'(z)\, p_1'(z) + \alpha[p_1''(z) + p_1'(z)^2] + \\ &\quad + \left[\frac{q(\alpha)}{q(\alpha+1)} - \frac{q(\alpha-1)}{q(\alpha)}\right] \exp(-p_1(z) = 0. \end{aligned} \tag{15}$$

On concludes from this equation that $q(\alpha)/q(\alpha+1)$ is a polynomial of at most the second degree in α, namely

$$-\frac{q(\alpha)}{q(\alpha+1)} = q_0 + s_1\alpha + r_2\alpha(\alpha+1) \tag{16}$$

with constants coefficients q_0, s_1, r_2. Inserting (16) into (15) then gives

$$p_0''(z) + p_0'(z)\, p_1'(z) = s_1 \exp[-p_1(z)], \tag{17}$$

$$p_1''(z) + p_1'(z)^2 = 2r_2 \exp[-p_1(z)]. \tag{18}$$

The solutions of these differential equations are

$$p_1 = \ln(r_0 + r_1 z + r_2 z^2), \qquad p_0' = \frac{s_0 + s_1 z}{r_0 + r_1 z + r_2 z^2} \tag{19}$$

with new constants r_0, r_1, s_0.

By use of (12), (14), (16), (19), the differential equation (9) becomes

$$\begin{aligned} &(r_0 + r_1 z + r_2 z^2)\, y''(z, \alpha) + [s_0 + s_1 z + (\alpha+1)(r_1 + 2r_2 z)]\, y'(z, \alpha) + \\ &\quad + [q_0 + s_1\alpha + r_2\alpha(\alpha+1)]\, y(z, \alpha) = 0 \end{aligned} \tag{20}$$

with arbitrary constants $r_0, r_1, r_2, s_0, s_1, q_0$. This is a differential equation of hypergeometric type. The same differential equation may be obtained from (8).

3. Discussion of the solution

While it is clear from (20) that the functions which satisfy the postulates mentioned in Section 2 are hypergeometric functions, including their special and limiting cases, it is worthwhile to give explicit expressions for $y(z, \alpha)$, $\Phi(z, \alpha)$, $Y(z, \alpha)$ in familiar notations, because, by TRUESDELL's results, to exhibit such expressions is tantamount to giving explicit generating functions and addition theorems for the various simple special functions.

We must distinguish several cases according to the character of the polynomial coefficient $r_0+r_1z+r_2z^2$. It can have two different roots, two equal roots, one single root ($r_2=0$), or none at all ($r_2=r_1=0$, $r_0\neq 0$).

In each case the procedure is to give a suitable solution of (20) in terms of the conventional notations of the simple special functions, to multiply it by a factor which depends only on α such that (4) is satisfied — this is then the $y(z, \alpha)$ — to evaluate $q(\alpha)$ from (16) and $\Phi(z, \alpha)$ from (13), and finally to write down the $Y(z, \alpha)$ according to (5).

Case 1. The polynomial $r_0+r_1z+r_2z^2$ has two distinct roots, which can be assumed to be at $z=0$ and $z=1$ (see the Lemma); that is, $r_0=0$, $r_1=-r_2$. Furthermore, one can assume $r_2=1$. For convenience s_0, s_1, q_0 are expressed by other constants a, b, c in the following way:

$$s_1 = a + b - 1, \qquad q_0 = a\,b, \qquad s_0 = 1 - c.$$

Then the differential equation (20) reads

$$\begin{aligned} z(z-1)\,y''(z,\alpha) + [(a+b+2\alpha+1)\,z - (c+\alpha)]\,y'(z,\alpha) + \\ + (a+\alpha)\,(b+\alpha)\,y(z,\alpha) = 0. \end{aligned} \tag{21}$$

This is the hypergeometric differential equation in normal form with the parameters $a+\alpha$, $b+\alpha$, $c+\alpha$.

From (16) one obtains

$$q(\alpha) = \frac{(-1)^\alpha}{\Gamma(a+\alpha)\,\Gamma(b+\alpha)}, \tag{22}$$

apart from a trivial constant factor, which can be omitted. (19) gives

$$\begin{aligned} p_1(z) &= \ln[z(z-1)], \\ \exp p_0(z) &= z^{c-1}\,(z-1)^{a+b-c}, \end{aligned} \tag{23}$$

again omitting a trivial constant factor. Inserting these results into (13) leads to

$$\Phi(\alpha) = \frac{1}{\Gamma(a+\alpha)\,\Gamma(b+\alpha)}\, z^{c+\alpha-1}(1-z)^{a+b+\alpha-c}, \tag{24}$$

up to a constant factor.

Not every solution of the differential equation (21) is a solution of the ascending F-equation (4). But from well known properties of the hypergeometric function, one verifies immediately that

$$y(z, \alpha) = \frac{\Gamma(a+\alpha)\,\Gamma(b+\alpha)}{\Gamma(c+\alpha)}\, {}_2F_1(a+\alpha, b+\alpha; c+\alpha; z) \tag{25}$$

has all the required properties.

Finally one obtains from (5) with $z=\xi$

$$Y(z,\alpha)=\frac{1}{\Gamma(c+\alpha)}\,z^{c+\alpha-1}(1-z)^{a+b+\alpha-c}\,{}_2F_1(a+\alpha,\,b+\alpha;\,c+\alpha;\,z). \tag{26}$$

A whole set of other solutions is obtained if one replaces the hypergeometric function $F(a+\alpha,\,b+\alpha;\,c+\alpha;\,z)$ by one of the functions

$$(1-z)^{-a-\alpha}\,{}_2F_1\left(a+\alpha,\,c-b;\,c+\alpha;\,\frac{z}{z-1}\right),$$

$$(1-z)^{c-a-b-\alpha}\,{}_2F_1(c-a,\,c-b;\,c+\alpha;\,z),$$

$$\Gamma(c+\alpha)\,\Gamma(c-a-b-\alpha)\,{}_2F_1(a+\alpha,\,b+\alpha;\,a+b+\alpha-c+1;\,1-z),$$

$$\frac{\Gamma(c+\alpha)}{\Gamma(b+\alpha)}\,(-1)^{\alpha}z^{-a-\alpha}\,{}_2F_1(a+\alpha;\,1-c+a;\,1-b+a;\,z^{-1}).$$

But, of course, all these solutions can be expressed in terms of any two which are linearly independent.

In each of these examples, the hypergeometric function ${}_2F_1(A,\,B;\,C;\,u)$ can be replaced by

$${}_2\Psi_1(A,\,B;\,C;\,u)=\frac{\Gamma(A')\,\Gamma(B')\,\Gamma(C-1)}{\Gamma(A)\,\Gamma(B)\,\Gamma(C'-1)}\,u^{C'-1}\,{}_2F_1(A',\,B';\,C';\,u)$$

with $A'=A-C+1$, $B'=B-C+1$; $C'=2-C$. Thus one obtains other forms of the solutions.

Case 2. The polynomial $r_0+r_1z+r_2z^2$ has two equal roots. Using the lemma, one can assume that $r_0=r_1=0$, $r_2=1$. We put $q_0=a(a-c+1)$, $s_1=2a-c$. Then the following results, which can be derived similarly as in Case 1, hold:

$$z^2\,y''(z,\alpha)+[s_0+(2a+2\alpha-c+2)\,z]\,y'(z,\alpha)+$$
$$+(a+\alpha)(a+\alpha-c+1)\,y(z,\alpha)=0,$$

$$y(z,\alpha)=(-1)^{\alpha}z^{-a-\alpha}\,\Gamma(a+\alpha)\,{}_1F_1\left(a+\alpha;\,c;\,\frac{s_0}{z}\right),$$

$$\Phi(z,\alpha)=\frac{(-1)^{\alpha}}{\Gamma(a+\alpha)\,\Gamma(a-c+\alpha+1)}\,e^{-\frac{s_0}{z}}\,z^{2a+2\alpha-c},$$

$$y(z,\alpha)=\frac{z^{a+\alpha-c}}{\Gamma(a+\alpha-c+1)}\,e^{-\frac{s_0}{z}}\,{}_1F_1\left(a+\alpha;\,c;\,\frac{s_0}{z}\right).$$

Instead of the confluent hypergeometric function, ${}_1F_1\left(a+\alpha;\,c;\,\frac{s_0}{z}\right)$, one can also take the function

$${}_1\Psi_1\left(a+\alpha;\,c;\,\frac{s_0}{z}\right)=\frac{\Gamma(a+\alpha-c+1)\,\Gamma(c-1)}{\Gamma(a+\alpha)\,\Gamma(1-c)}\left(\frac{z}{s_0}\right)^{c-1}{}_1F_1\left(a+\alpha-c+1;\,2-c;\,\frac{s_0}{z}\right).$$

A simple special case is $s_0=0$, which however, will not be discussed further.

Case 3. The polynomial $r_0+r_1z+r_2z^2$ has only one single root. One therefore assumes $r_2=0$, $r_1=1$, $r_0=0$. Here we must distinguish between $s_1\neq 0$ and $s_1=0$.

Subcase 3a. $s_1 \neq 0$. Then one can put $s_1 = -1$ without loss of generality. Introducing the constants $c = s_0 + 1$, $a = -q_0$, we obtain

$$z\,y''(z,\alpha) + (c+\alpha-z)\,y'(z,\alpha) - (a+\alpha)\,y(z,\alpha) = 0,$$

$$y(z,\alpha) = \frac{\Gamma(a+\alpha)}{\Gamma(c+\alpha)}\,{}_1F_1(a+\alpha;\,c+\alpha;\,z),$$

$$\Phi(z,\alpha) = \frac{1}{\Gamma(a+\alpha)}\,e^{-z}\,z^{c+\alpha-1},$$

$$y(z,\alpha) = \frac{1}{\Gamma(c+\alpha)}\,e^{-z}\,z^{c+\alpha-1}\,{}_1F_1(a+\alpha;\,c+\alpha;\,z).$$

A further solution is obtained with the same replacement of the confluent hypergeometric function ${}_1F_1$ by ${}_1\Psi_1$, as given in Case 2.

Subcase 3b. Let $s_1 = 0$. Then

$$z\,y''(z,\alpha) + (s_0+\alpha+1)\,y'(z,\alpha) + q_0\,y(z,\alpha) = 0,$$

$$y(z,\alpha) = (-1)^{\alpha}\,q_0^{\alpha/2}\,z^{-(s_0+\alpha)/2}\,J_{s_0+\alpha}\left(\sqrt{4q_0 z}\right),$$

$$\Phi(z,\alpha) = (-q_0)^{-\alpha}\,z^{s_0+\alpha},$$

$$Y(z,\alpha) = q_0^{-\alpha/2}\,z^{(s_0+\alpha)/2}\,J_{s_0+\alpha}\left(\sqrt{4q_0 z}\right).$$

The Bessel function in these formulas can be replaced by the Neumann function or by one of the Hankel functions with the same argument and order.

Case 4. The polynomial $r_0 + r_1 z + r_2 z^2$ reduces to a constant, which is assumed to be 1. Then we again have to distinguish between the cases $s_1 \neq 0$ and $s_1 = 0$.

Subcase 4a. Let $s_1 \neq 0$. Without loss of generality we assume $s_1 = 1$, $s_0 = 0$ and put $q_0 = 1 + a$. Then

$$y''(z,\alpha) + z\,y'(z,\alpha) + (1+a+\alpha)\,y(z,\alpha) = 0,$$

$$y(z,\alpha) = (-1)^{\alpha}\exp\left(-\frac{z^2}{4}\right)D_{a+\alpha}(z),$$

$$\Phi(z,\alpha) = \frac{(-1)^{\alpha}}{\Gamma(a+\alpha+1)}\exp\left(\frac{z^2}{2}\right),$$

$$Y(z,\alpha) = \frac{1}{\Gamma(a+\alpha+1)}\exp\left(\frac{z^2}{4}\right)D_{a+\alpha}(z).$$

Instead of the function of the parabolic cylinder $D_{a+\alpha}(z)$, one can also take the following function:

$$\Gamma(a+\alpha+1)\,i^{\alpha}\,D_{-a-\alpha-1}(i\,z).$$

Subcase 4b. Let $s_1 = 0$. In this case the differential equation (20) become independent of α. Considering all possible values of s_0 and q_0, one arrives at the following pair of solutions ($a \neq 0$):

$$Y(z,\alpha) = a^{\alpha}\,e^{a\,z}, \qquad \tilde{Y}(z,\alpha) = a^{\alpha}(a\,z+\alpha)\,e^{a\,z},$$

$$\Phi(z,\alpha) = (-1)^{\alpha}\,a^{-2\alpha}\,e^{-2a\,z},$$

$$y(z,\alpha) = (-1)^{\alpha}\,a^{-\alpha}\,e^{-a\,z}, \qquad \tilde{y}(z,\alpha) = (-1)^{\alpha}\,a^{-\alpha}(a\,z+\alpha)\,e^{-a\,z}.$$

4. Remarks and discussion

In the foregoing results a constant factor (constant with respect to both variables z and α) can always be added in $y(z, \alpha)$ as well as in $Y(z, \alpha)$. Furthermore, the results remain correct if $y(z, \alpha)$ and $Y(z, \alpha)$ are multiplied by a periodic function of α with period 1. One also sees immediately that $y(z, \alpha)$, $\Phi(z, \alpha)$, $Y(z, \alpha)$ can in turn be changed into $Y(z, -\alpha)$, $\Phi(z, -\alpha)^{-1}$, $y(z, -\alpha)$ to give a solution with the properties mentioned in Section 1. Although one does not obtain new solutions in this way, the totality of solutions having been exhausted by the cases 1 to 4 to within linear transformations of the independent variable, it may, however, be convenient to put the solutions into a different form in this way.

Special situations, such as terminating sequences $y(z, \alpha)$, are not explicitly treated; it is easy to make the necessary modifications in every case.

Of course, the limiting cases 2 to 4 can also be obtained by going to the respective limits not in the differential equations, but in the solutions themselves, $y(z, \alpha)$, $\Phi(z, \alpha)$, $Y(z, \alpha)$, of Case 1. Therefore everything is theoretically contained in Case 1, explicitly in the formulas (24), (25), (26) and in those which arise by replacing α by $-\alpha$ and y, Φ, Y by Y, Φ^{-1}, y.

We repeat that the solutions $y(z, \alpha)$ of the ascending F-equation written down explicitly in the Cases 1 to 4 satisfy the functional equation (7) with $dz/d\xi = 1$, and furthermore we remark that (7) and (4) can be combined into the difference equation or the recurrence relation

$$y(z, \alpha + 1) + \frac{\Phi'(z, \alpha)}{\Phi(z, \alpha)} y(z, \alpha) - \frac{\Phi(z, \alpha - 1)}{\Phi(z, \alpha)} y(z, \alpha - 1) = 0.$$

Legendre functions are special cases of hypergeometric functions and can be expressed in terms of these in various ways. This gives rise to many functions $y(z, \alpha)$ that can be expressed by Legendre functions. However, we shall not go into the details here.

Queen's College, Queens, N.Y.
and
Technische Hochschule Aachen

(Received June 26, 1959)

EDITORIAL BOARD

CONTENTS

Druck der Universitätsdruckerei H. Stürtz AG., Würzburg
Printed in Germany